

Durham Research Online

Deposited in DRO:

29 July 2016

Version of attached file:

Accepted Version

Peer-review status of attached file:

Unknown

Citation for published item:

Menzies, V. and Kasim, A. and Kokotsaki, D. and Hewitt, C. and Akhter, N. and Collyer, C. and Younger, K. and Wiggins, A. and Torgerson, C. (2016) 'Hallé SHINE on Manchester : evaluation report and executive summary.', Project Report. Education Endowment Foundation.

Further information on publisher's website:

<https://educationendowmentfoundation.org.uk/evaluation/projects/shine-on-manchester/>

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.



Hallé SHINE on Manchester

Evaluation Report and Executive Summary

January 2016

Independent evaluators:



Victoria Menzies, Adetayo Kasim, Dimitra Kokotsaki, Catherine Hewitt (University of York), Nasima Akhter, Clare Collyer, Kirsty Younger, Andy Wiggins and Carole Torgerson

Contents

List of Tables	3
List of Figures.....	4
Executive Summary	5
Introduction.....	7
Methods.....	14
Impact Evaluation	29
Process Evaluation	55
Conclusion	69
References	710
Appendices	73

List of Tables

Table 1. Executive Summary Table.....	6
Table 2. Trial design	14
Table 3. Intervention attendance by nominated pupils	25
Table 4. Process Evaluation Summary.....	26
Table 5. Timeline	27
Table 6. First Year (Pilot) School Level Baseline Characteristics (data from 2012).....	33
Table 7. Description of pupil level characteristics (first year)	34
Table 8. Second year school level characteristics (data retrieved from Ofsted data dashboard (Ofsted, c2012-2016) for 2013)	35
Table 9. Description of pupils level characteristics (second year)	36
Table 10. Third year school level characteristics (data retrieved from Ofsted data dashboard (Ofsted, c2012-2016) for 2014)	37
Table 11. Description of pupil level characteristics (third year)	38
Table 12. Pooled effect size and estimate together with 95% confidence intervals for the primary outcome (reading)	39
Table 13. Pooled effect size from intention-to-treat analysis of the secondary outcomes using multi-level model	40
Table 14. Pooled effect size and estimate together with 95% confidence intervals for the primary outcome (reading)	41
Table 15. Pooled effect size from intention-to treat analysis of the secondary outcomes using multi-level model	42
Table 16. Minimum detectable effect size for second year trial	42
Table 17. Pooled effect size together with 95% confidence intervals for the primary outcome (reading) for year 3 trial.....	43
Table 18. Effect size and estimates from intention-to-treat analysis of the secondary outcomes using multi-level model.....	44
Table 19. Minimum detectable effect size for third year trial.....	44
Table 20. Analysis of primary outcome by year group	45
Table 21. Subgroup analysis of primary outcome for FSM pupils.....	47
Table 22. Average compliance in intervention and control schools, complier (ϕ) was the difference in average compliance between the intervention and the control groups.	49
Table 23. Costs per pupil for the Hallé SHINE on Manchester intervention	54

List of Figures

Figure 1. Theory of change for <i>HSoM</i> programme from 'Making it work' handbook (p.3)	9
Figure 2. First Year (Pilot Trial) participant flow diagram	30
Figure 3. Second Year Participant Flow Diagram	31
Figure 4. Trial Third Year Participant Flow Diagram	32
Figure 5. Analysis of primary outcome by year group	46
Figure 6. Exploratory analysis of intervention effect on KS2 reading	48
Figure 7. Exploratory analysis of intervention effect on KS2 maths	48
Figure 8. CACE analysis based on asymptotic 95% confidence intervals for pupils with more than 50% compliance	50
Figure 9. CACE analysis for Year 5 from the pilot trial based on non-parametric bootstrap confidence intervals and a grid of compliance thresholds	51
Figure 10. CACE analysis for Year 6 from the second year trial based on non-parametric bootstrap confidence intervals and a grid of compliance thresholds	52
Figure 11. CACE analysis for Year 4 from the third year trial based on non-parametric bootstrap confidence intervals and a grid of compliance thresholds	53
Figure 12. Budget Template for Intervention taken from Hallé SHINE on Manchester: Making it work (SHINE, 2012)	54

Executive Summary

The project

The *Hallé SHINE on Manchester (HSoM)* programme is a Saturday school educational programme designed to increase the reading and maths attainment, as well as engagement with school, of underachieving and disadvantaged pupils at Key Stage 2.

Developed in collaboration between the SHINE Trust and *Hallé* Orchestra, the intervention provides additional school-based literacy and numeracy lessons, based on musical themes, as well as visits to *Hallé* rehearsals, performances and other theme-based activities. 25 Saturday sessions, each lasting five hours, were planned for the intervention over the course of an academic year, delivered by qualified teachers, teaching assistants, peer mentors and professional musicians.

The evaluation consisted of two randomised controlled trials (RCTs), a pilot trial and a main trial, and a process evaluation conducted with primary schools in the Manchester area between July 2012 and July 2015. The pilot trial involved 361 Year 5 and 6 pupils in 18 schools; the main trial involved 2,306 Year 4, 5 and 6 pupils in 38 schools. The second year involved both the pilot and main trial running concurrently. The third year involved main trial schools only.

Key Conclusions

Key conclusions

1. There is no evidence that the *Hallé Shine on Manchester* programme had an impact on the reading attainment of children in the trial. This finding was consistent across all age groups, for pupils eligible for free school meals, and for all three years of the trials.
2. There is no evidence that the *Hallé Shine on Manchester* programme had an impact on the attainment of children in mathematics, or attitudes to reading, maths, music and school of the children in the trial.
3. Attendance of eligible pupils was often low and considered as a barrier to successful implementation. Reasons for low attendance included pupils' lack of availability to attend the Saturday sessions, variable parental engagement with the programme and limited time at the beginning of the programme for schools to engage children and parents.
4. The process evaluation revealed a positive picture of involvement and engagement for those pupils who attended the Saturday school activities. Evaluators observed good working relationships between the teachers and pupils and positive and purposeful learning environments in lessons. All stakeholders felt pupils were making noticeable improvements in behaviour, confidence and the development of social skills.

Security rating

Findings from this study have moderate security. The study used randomised controlled trial designs (RCTs) with random allocation at the level of the year group (i.e., schools identified eligible pupils from two year groups, one of which was randomly selected to participate in the intervention). The trials were efficacy trials; efficacy trials aim to test whether the intervention can succeed under ideal conditions.

The trials were smaller in size than expected because not as many pupils were recruited as planned and because a significant proportion of pupils did not complete all the tests at the end of the project (15% in Year 1, 22% in Year 2 and 25% in Year 3).

Results

The findings demonstrated no evidence that the *Hallé Shine on Manchester* programme had an impact on the primary outcome of reading attainment of children in the trial. This outcome was consistent across all age groups, for both the pilot and main trial. This effect also applied to pupils eligible for free school meals.

No positive effects were observed for the secondary outcome measures of attainment in maths and attitudes to reading, maths and music. Further exploratory analysis of Key Stage 2 results also showed no signs of positive impacts in reading, maths and writing for Year 5 and 6 pupils in the first two years of the trial.

Pupil recruitment and attendance was often low and a barrier to successful implementation. Reasons for low attendance included pupils' lack of availability to attend the Saturday sessions, variable parental engagement with the programme and limited time at the beginning of the programme for schools to engage children and parents. Nevertheless, further analysis that took attendance into account did not alter the overall finding.

In contrast, the process evaluation revealed a positive picture of involvement and engagement for those pupils who attended Saturday school activities. Good working relationships between the teachers and pupils were observed, with positive and purposeful learning environments demonstrated in lessons. Teachers, parents and musicians all felt pupils were making noticeable improvements in behaviour, confidence and the development of social skills. Although pupil engagement and participation in lessons were generally very good, more challenging feedback may have helped pupils' learning.

These results are generally consistent with existing evaluations of out-of-school programmes, showing small or no effects on literacy or maths achievement when robust research methods are adopted. Small positive effects on academic outcomes have previously been observed in other research when qualified teachers were used (as is the case in this evaluation); however, this study did not confirm this finding. Another EEF funded pilot evaluation of a similar *SHINE* programme with pupils in Year 7 also found only very small non-statistically significant effects of the programme on numeracy and literacy (Menzies et al., 2015).

How much does it cost?

The costs of the intervention were calculated by dividing the grant provided to each school to run the programme for a year by the number of pupils the intervention could accommodate. This cost estimate is £727 per pupil per year.

Table 1. Executive Summary Table

Group	Effect Size (95% confidence interval)	<i>Estimated months' progress</i>	<i>Security rating</i>	<i>Cost</i>
<i>Hallé SHINE on Manchester vs control (Year 1)</i>	0.03 (-0.27, 0.34)	1 months	3	££££
<i>Hallé SHINE on Manchester vs control (Year 2)</i>	-0.10 (-0.26, 0.07)	-2 months	3	££££
<i>Hallé SHINE on Manchester vs control (Year 3)</i>	0.10 (-0.20, 0.40)	2 months	3	££££

Introduction

Intervention

Overview

The *Hallé SHINE on Manchester (HSoM)* programme is a Saturday school educational programme developed by the *SHINE* Trust (known hereafter as *SHINE*) in conjunction with the Hallé Orchestra's education team. By providing additional tuition in English and mathematics in a form different from weekday schooling (smaller class sizes with more support and theme-based, interactive lessons linked to music creation and performance), the programme aims to increase the reading and maths attainment, as well as engagement with school, of underachieving and disadvantaged pupils. The *Hallé SHINE on Manchester* programme is based on the existing *SHINE on Saturday* programme which has been running in schools in London since 2001. The key development for this project is the collaboration with the Hallé: all lessons including literacy and numeracy are designed around a music theme.

The *HSoM* programme involves weekly Saturday school sessions throughout the school year delivered by primary schools in the Manchester area. The programme is targeted at disadvantaged and/or underachieving Key Stage 2 pupils in a 'host' school and their local partner schools. The format of the programme is specified by *SHINE* and the Hallé's education team and involves numeracy, literacy and music workshops built around termly music themed projects supplemented by visits to Hallé rehearsals and performances, as well as other theme-based activities and trips. Lesson content and programme detail are planned and delivered by the project staff, some of whom are local teachers, at the host school.

The HSoM programme in detail

For this evaluation host schools were funded by a grant from *SHINE* to run the *Hallé SHINE on Manchester* programme for pupils in years 4, 5 and 6 (depending on the year of the trial and on allocation to intervention or control condition, see below).

The host school intervention delivery projects (hereafter named 'projects') were designed to accommodate around 60 pupils in total drawn from the host school and local partner schools. All host schools were provided with the *Hallé SHINE on Manchester* 'Making it Work' handbook to guide them in setting up and running the intervention. This includes guidance on the recruitment of staff required for the project and their initial training, premises management, budgeting and finance, selecting eligible students, encouraging attendance, *SHINE*'s monitoring and evaluation arrangements and guidance on the curriculum.

The programme was designed to run for 25 Saturdays during term time of each school year with each session lasting five hours. Most Saturday sessions were expected to consist of three different lessons/activities over the day (including numeracy and literacy lessons) with lunch provided by the project, as well as break time and some time when the whole group was together. A core part of the programme is the high staff-to-pupil ratio (1:15 teachers, or 1:9 including teaching assistants). Each project was staffed by: four qualified teachers (the project manager and three tutors); three qualified teaching assistants; and three peer mentors (pupils from local secondary schools who were paid for their time in gift vouchers). All were recruited and employed by the host schools following guidance provided

by *SHINE*. Two Hallé musicians were also assigned to each project and they attended for two hours for 20/25 sessions in each project in order to deliver the creative music programme. Pupils were divided into three classes of up to 20 students which included pupils from both year groups.

Each project was responsible for developing their own curriculum to suit the needs of their students based on the guiding principles specified by *SHINE* in the 'Making it Work' handbook. These include:

- A core focus on literacy and numeracy, based on a musical theme set by the Hallé;
- Opportunities to revisit areas where students were struggling and to extend into areas outside the weekday subjects;
- Project work, hands-on learning and opportunities for enrichment which enabled students to apply their learning to real world contexts;
- A thematic approach which offered opportunities to explore new areas beyond the national curriculum;
- Flexibility of the curriculum to do project work, trips and visits which offer opportunities for children who were de-motivated or struggling with learning to re-engage;
- Emphasis on investigative and collaborative learning and research projects which encouraged problem-solving and independent learning.

The curriculum for each term was planned around a musical theme specified by the Hallé. The *HSoM* Programme Manager provided projects with a support pack containing information about music and the composer, ideas for lessons as well as ideas for musical and creative enrichment activities to go with each musical theme. This theme formed the basis of the creative music project that the children developed over the course of each term with the two Hallé musicians.

Projects were expected to include appropriate educational visits and were given a budget to incorporate these. Some of the visits were already planned into the programme by the Hallé and included trips to Hallé rehearsals or concerts. Projects were free to plan other visits to complement the term's theme and connect to a specific subject area e.g., a trip to the space centre as part of 'The Planets' theme.

Each project manager was expected to complete a medium term plan for each term which demonstrated how learning linked up between Saturdays and how enrichment opportunities could be integrated. This plan was shared with the *HSoM* programme manager and with the tutors at that project. Tutors were then responsible for planning individual lessons each week. The project manager worked with tutors to ensure lessons were differentiated effectively and was responsible for quality assurance. The *HSoM* programme manager maintained regular contact with the project managers throughout the year and visited each project two or three times.

The 'Making it Work' handbook describes a clear process for maintaining pupil attendance and following up non-attendance at the programme for pupils who sign up to the programme. Poor attendance at projects was an issue throughout the evaluation which will be highlighted later in the results and process evaluation.

Theory of change

Before the start of the evaluation *SHINE* developed a theory of change model of how they felt different aspects of the *Hallé SHINE on Manchester* programme may impact on participants (SHINE, 2012, p.3 reproduced below in Figure 1).

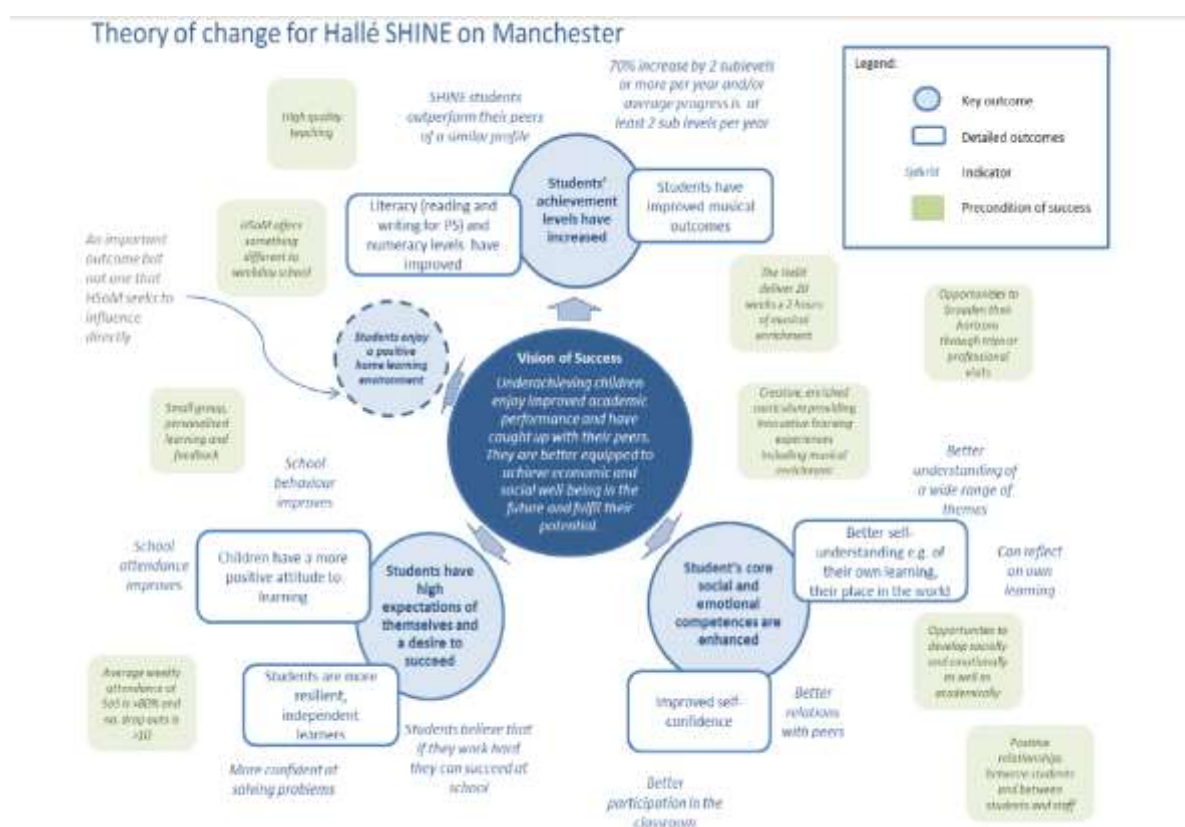


Figure 1. Theory of change for *HSoM* programme from 'Making it work' handbook (p.3)

Background evidence

Existing evidence for the intervention

There have been few robust research studies of Saturday school programmes. The most relevant evaluation studies described in more detail below have been of out-of-school (which include before and after school programmes as well as weekend programmes) or summer school programmes mostly delivered in the United States (US). The results of these studies have generally been mixed, possibly due to the variety of different kinds of programmes evaluated, the quality of programme delivery and variations in the evaluation designs (see Kidron and Lindsay, 2014; Lauer et al, 2016; Crawford, 2011; Fashola, 1998).

A high quality recent meta-analytic review by Kidron and Lindsay (2014) included 30 rigorous evaluation studies of increased learning time programmes conducted in the US since 1998. The authors looked at the impact of different types of increased learning time programmes, of different aspects of programmes and of the effects of the programmes on different subgroups. They reported that out-of-school programmes had a very small statistically significant effect on academic motivation (effect size of 0.04); however, they found no evidence of effect of these programmes on literacy or mathematics achievement. When certified teachers were employed in delivering any increased learning time programme (not just out of school programmes), as is the case for *HSoM*, the authors found a small statistically significant effect on literacy and maths achievement (effect sizes of 0.18 and 0.09 respectively). They also found traditional teaching during the programmes to have small statistically significant effects (effect size of 0.14 for literacy achievement and 0.10 for maths achievement). In terms of the groups of students supported by these programmes, the authors found students performing below standards to benefit in literacy achievement (effect size of 0.56). They also found a small statistically significant impact on literacy achievement and maths achievement for 'not at risk' students (effect sizes of 0.10 and 0.09 respectively). There was no impact of the programmes on students from low-income households on any aspect. The authors found a statistically significant negative impact of programmes on middle grade students (aged 11 to 14) literacy achievement (effect size of -0.21), but a very small statistically significant positive impact of programmes on elementary school students (aged 5 to 11) (effect size 0.07). It is worth noting that only one of the studies included in this meta-analysis was of a weekend-based out-of-school programme which focused on a hands-on science and technology curriculum over 8 Saturday sessions and which found no statistically significant effect on the short-term outcomes measures (Martinez & Cosentino de Cohen, 2010). However, 21 of the 30 studies included in the meta-analysis were of optional out-of-school-time programmes (either before or after school programmes).

Another meta-analysis of the effects of out-of-school-time programmes for at risk students in the US, relevant to the disadvantaged target group of EEF funded projects, (Lauer et al., 2006) found small but significant positive effects on both reading and maths achievement (0.05-0.13 and 0.09-0.17). This review identified 35 studies (including quasi-experimental studies); however, only one of these studies included a Saturday school (Harlow & Baenen (2001), and this was in combination with a summer school - the others were after-school programmes or summer schools.

A meta-analysis conducted by Crawford (2011) looking at evaluations of US out-of-school programmes with a mathematics and/or literacy focus, found an overall effect size of 0.40 for impact on reading and maths outcomes. This included 23 effect sizes from studies of students between Kindergarten and Grade 8 involved in studies between 2000 and 2009. However the inclusion criteria for this project did not include a judgement about the quality of the research design and any study that included an effect size (or allowed an effect size to be calculated) was included. It therefore is likely that the effect size has been inflated by the inclusion of results from less rigorous studies.

The studies above all focused on US-based programmes and evaluations which may not be generalizable to the English education system or population. There is no rigorous research focusing on the impact of out-of-school programmes in the United Kingdom (UK). MacBeth et al. (2001) looked at the impact of out-of-school-hours programmes on secondary school students in the UK using a longitudinal cohort design. They tracked the progress of students in two year-groups (Year 7 and Year 9) in 52 schools where out-of-school programmes were being offered. The study suggested that out of hours study support had a positive impact on results (half a grade in maths and English GCSEs and a third of a grade in maths and three quarters of a grade in science at Key Stage 3) compared to what would have been expected from baseline measures. Those pupils who attended out-of-school programmes also had

better school attendance and more positive attitudes. This study has a limited research design, however, without a randomised comparison group meaning that it is impossible to control for any unobservable characteristic including motivation in attending the programme and conclusions are therefore tentative.

Because of the musical projects, workshops, theming and involvement of the Hallé in the *HSoM* programme we should also consider the background evidence for music education and enrichment programmes. A meta-analysis by Standley (2008) looked at the impact of music education on reading achievement and found modest effects. However, these effects were largest in studies at lower ages (4 - 6 year olds) and declined as children got older (elementary school and beyond). On the other hand a recent EEF evaluation conducted by Haywood et al. (2015) found that a music workshop had no impact on maths or literacy achievement for Year 2 pupils compared to a drama workshop.

An in-school music programme by the New London Orchestra with similar aims to the *HSoM* programme was evaluated through a small study with six and seven year olds, by Welch et al. (2012). This used a non-randomised pre- post design (intervention n=209, control n=61). The results showed a positive effect (intervention gain 8.4 months compared to 1.8 months for the controls). However, these results should be viewed with caution, due to a number of factors, including design limitations of the study.

A report using *SHINE*'s own data by the National Foundation for Educational Research (NFER) and commissioned by *SHINE* found that students attending *SHINE on Saturday* schools (the existing programme on which *HSoM* is based) made the expected progress at Key Stage 2 (Chamberlain et al., 2011). Perceptions about confidence, self-esteem and attitudes to learning were found to have improved as evidenced by interviews with students and teachers involved.

Most research into the effectiveness of out-of-school programmes has not used rigorous research designs (e.g. Scott-Little, Hamann & Jurs, 2002; Fashola, 1998). The non-mandatory nature of these programmes which happen, by definition, in students' free time, means that there is usually some uncontrolled factor that influences why some children attend the programme and why some do not (Fashola, 1998). This is rarely accounted for in the designs of research looking at the programmes' effectiveness. We believe that this study provides a non-biased way of assessing the impact of the *Hallé SHINE on Manchester* programme by using randomisation to assign pupils to be invited to attend or not. However, this also means that the effect of attending the programme may be understated due to including pupils who did not attend the intervention in the (intention to treat) analyses (which is the correct analytical approach).

Stage of development

The *HSoM* programme is based on the existing *SHINE on Saturday* programme which has been delivered in schools in London over the last 12 years and is currently attended by over 1000 pupils in more than 60 schools. In this programme schools are provided with a manual detailing the running of their project which includes staffing, the format of the day, the planning and the monitoring of programmes, as well as on-going support from *SHINE*. For *HSoM* the manual was adapted to reflect the involvement of the Hallé and the music elements of the programme. In addition schools receive other resources for the Hallé musician-led creative music workshops.

As *HSoM* is a new programme this evaluation is an efficacy trial to test the promise of the intervention.

Rationale

HSoM is a developed, replicable intervention ready for evaluation. The rationale for evaluation is based on the limited empirical evidence demonstrating promise of the intervention and its potential application to schools serving disadvantaged populations. An independent evaluation using rigorous design and methods was, therefore, timely. The evaluation focus was on establishing an unbiased estimate of impact of the intervention on short-term academic outcomes (literacy and numeracy) and attitudinal outcomes compared with an untreated, business-as-usual control group.

Evaluation objectives

The independent evaluation aimed to address the following research questions:

1. What is the impact of the *Hallé SHINE on Manchester* Saturday school programme on academic attainment outcomes at post-test, specifically literacy and numeracy? (pilot and main trial)
2. What is the impact of the *Hallé SHINE* programme on attitudes to school and learning, KS2 results and school attendance at post-test? (pilot and main trial)
3. What is the impact of the *Hallé SHINE* programme on music outcomes, specifically pupil attitudes towards music and self-assessed music ability at post-test? (main trial only)
4. What are the management and organisational issues associated with implementing the intervention? (main trial only)

Project team

SHINE: responsible for programme development, recruitment to the project, awarding of grants, and delivery of intervention

Paul Carbury – *SHINE* Chief Executive
Fiona Spellman – *SHINE* Senior Programme Manager
Caroline Davies – *SHINE* Senior Programme Manager

Hallé: responsible for recruitment to the project, development of project resources and musical themes, monitoring delivery of project and supporting schools

Steve Pickett – Hallé Education Director
Carolyn Davis – Hallé *SHINE on Manchester* Programme Manager (until July 2014)
Jo Brockbank – Hallé *SHINE on Manchester* Programme Manager (from Sept 2014)
Hallé musicians - responsible for delivery of creative music aspects of the programme (2 per project each term)

Durham University and University of York: responsible for evaluation

Durham University, School of Education

Professor Carole Torgerson – Principal Investigator

Durham University, CEM

Dr Andy Wiggins – CEM Project Manager

Victoria Menzies – Trial Coordinator

Clare Collyer – Project Administrator (main trial)

Kirsty Younger – Project Administrator (pilot)

Dr Dimitra Kokotsaki – Project Researcher (process evaluation)

Durham University, Wolfson Research Institute for Health and Wellbeing

Dr Adetayo Kasim – Project Statistician

Dr Nasima Akhter – Project Statistician (third year)

University of York, York Trials Unit

Dr Catherine Hewitt – Senior Statistician

Ethical review

Ethical approval was obtained from the Board of Ethics in the School of Education, Durham University on the 22nd August 2012. Consent was given by the head teacher of each school and all schools signed a memorandum of understanding agreeing to the requirements of the trial. Pre- and post-testing was conducted by schools across the whole year groups involved rather than just the eligible pupils to avoid singling out disadvantaged students. Opt-out parental consent for data usage was obtained by the evaluators. For children attending the intervention, the school obtained parental opt-in consent.

Trial registration

The trial was registered with ISCRTN in September 2013 (ISCRTNN99779220).

Methods

Trial design

The design of the evaluation is a series of two-armed pragmatic cluster randomised controlled trials (RCT), with random allocation at the level of the year group. The pragmatic design reflected as closely as possible the implementation of the programme in ‘real life’. Teachers chose eligible pupils for the study as they would do in ‘normal’ implementation. This enables the results to be generalizable to similar pupils and schools outside the trial. Thus, each school provided both an intervention year group (eligible pupils invited to attend the *HSoM* programme) and a control year group (business as usual: no invitation to *HSoM*) determined through random allocation. This design was developed with input from *HSoM* and EEF as it was deemed to be more acceptable to schools than individual pupil randomisation. In addition, the group (or cluster) randomisation design enabled potential contamination at the level of class to be minimised as whole year groups (or classes) were randomised. The evaluation ran for three years with a separate RCT, involving different pupils, taking place each year. Schools were recruited to take part in the evaluation for two years, with pilot schools starting in the project’s first year (hereafter ‘first year’) and main trial schools starting in the project’s second year and continuing into the third year (hereafter ‘second’ and ‘third’ years). For the first year of each school’s involvement Year 5 and Year 6 pupils participated in the study, while for the second year of involvement Year 4 and Year 5 pupils participated (see Table 2 below).

Table 2. Trial design

Trial Year	Pilot Schools		Main Trial Schools	
	Year groups involved	Number of schools (number of projects)	Year groups involved	Number of schools (number of projects)
1	Year 5* & 6*	14 (3)	-	-
2	Year 4 & 5*	18 (3)	Year 5* & 6*	20 (5)
3	-		Year 4 & 5	19 (5)

*Pupils for whom Key Stage 2 results were available.

At the start of each year, teachers nominated eligible pupils from both participating year groups and then concealed and independent random allocation (by the trial statistician – AK) assigned one of the participating year groups to intervention and one to control within each school.

Eligibility criteria for pupil nomination across the project was specified by *SHINE* as part of the intervention; nominated pupils needed to fit at least one criteria: underachievement, an indicator of disadvantage; English as an additional language or special educational needs.

The three years of the RCTs were analysed separately by year and by year-group and the results for each year were pooled across its year-group in a meta-analysis, as pre-specified in the trial protocol. The first year of the trial was run as an internal pilot trial where all processes related to the trial were tested and the acceptability and processes surrounding the delivery of the intervention were monitored and fed back into the design of the main trial. Note that the second year trial comprised of both the schools recruited during the pilot study and those recruited in the second year of the trial.

Intervention group allocation

Within each school, all children from the randomly allocated intervention year group, who had been nominated by teachers as eligible for the intervention (see Participant Selection section), were invited to sign up for and attend the *HSoM* programme by the schools involved. This was done through promotional material sent home with eligible pupils, school events and meetings with parents of eligible pupils. All eligible pupils were encouraged to attend by the school.

Control group allocation

The nominated pupils in the control year group within each school were not invited to attend the intervention and schools were asked not to inform pupils that they were part of the control group. Control group pupils in host schools may have been aware of the project due to displays in the school and staff working at the project also teaching in the school, but should not have been exposed to the intervention. There were a few deviations to protocol where control pupils were invited to attend the intervention, see below. Control group pupils may have chosen to attend other Saturday programmes; however, this was not monitored. The comparison group was therefore considered as a 'business as usual' control group.

A pragmatic randomised trial is the most rigorous evaluative design. It avoids selection bias, because random allocation determines which groups are in the intervention group and which groups are in the control group. Combining three smaller trials conducted over the three years of the project increased the power of the evaluation compared to looking at the results for each of the individual years of the study. The decision to randomise at the year group level was more acceptable to *SHINE* and the schools involved, as every school was able to access the intervention. Due to the intervention taking place outside of school time it was expected that contamination to the control group would be minimal. The trial was designed, conducted and reported to CONSORT standards (Schultz, 2010) in order to minimise all potential threats to internal validity, such as selection bias and a range of post randomisation biases (Cook and Campbell, 1979; Shadish, Cook and Campbell, 2001; Torgerson and Torgerson, 2008). In this way unbiased estimates of impact of the intervention can be provided.

Outcome measures

The primary outcome of the trial was reading ability at the end of the school year that the intervention took place. This was measured by InCAS (Interactive Computerised Assessment System). Secondary outcomes were maths ability and attitudes to school, maths and reading scales measured by InCAS as well as a music attitudes and ability scale also measured at the end of the school year.

Academic outcomes were measured using the InCAS assessment provided by the Centre for Evaluation and Monitoring (Centre for Evaluation and Monitoring, 2016). The assessment is an adaptive computerised assessment suitable for children in Year 2 to Year 6. It contains multiple modules of maths and literacy assessment as well as a module looking at attitudes to school and learning. Children sit the assessments on their own at a computer while wearing headphones; the instructions and questions are read out aloud by the software and each child completes the assessment at their own pace. The adaptive nature of the test means that the items that are presented in the assessment are dependent on the child's age and their responses to previous questions. No adult input should be needed during the assessment. In most instances the delivery of assessments was done by schools' personnel

and data were returned automatically to CEM for processing. In one school, assessments were delivered by researchers from the evaluation team when the school's IT equipment had been stolen. The marking of the assessments was done by the software. Although the testers were not blinded, the nature of the computer adaptive tests limits the potential for bias to be introduced. The technical guide for this assessment reports predictive validity using correlations between InCAS modules scores in Year 3 with the PIPS Year 4 scores one year later (0.72 for reading predicting reading, 0.70 for maths predicting maths). Rasch person reliability for all modules ranges from 0.89 and 0.97.

The InCAS assessment was chosen due to the adaptive nature of the test which gives a more accurate reflection of a student's ability across the range of possible abilities. It also allowed the test to be delivered by schools without testers having to visit schools, while still minimising testing bias. The intervention aimed to improve core academic outcomes so literacy and maths outcomes were selected. A discussion with the developer led to reading being chosen as the primary outcome and the other InCAS modules as secondary outcomes.

Music outcomes were measured at the end of the second and third years via a paper survey delivered by the class teacher and these were returned to the research team for processing. The survey took established items from one of the research team's previous work (Kokotsaki, 2015; Kokotsaki, 2016) which looked at attitudes to music and self-appraisal of musical skills. The *Liking Music* part of the survey was adapted from the attitudes to music scale ($\alpha=0.87$) and the *Making and Understanding* part of the survey was adapted from the self-assessment rating scale ($\alpha=0.83$) that were previously used in a Nuffield-funded project which explored the transition in music education from the primary to the secondary school.

Additional measures (not pre-specified in the protocol) of fine grade average point scores for Key Stage assessment and school absence (defined as percentage of possible sessions a pupil was absent for) were collected through a request to the National Pupil Database. Key Stage 2 results were only available for pupils who had completed year 6 by July 2015 (indicated by * in Table 2). Absence data were only available for the first two years of the trial. Data were requested from the National Pupil Database in October 2015 for pupils in all three years of the trial.

Assessment delivery

Pre-test assessments were completed in September/October each year and post-testing was completed in June/July the following year. In the first year, pupils completed the Reading, General Maths, Mental Arithmetic, and Attitudes InCAS assessments. Schools were originally asked to also complete the Spelling assessment, but due to difficulties with time for completing so many assessments we removed the spelling module from the trial requirements. In the second and third years pupils also completed the paper music survey in June/July. All pupils in the year groups involved (not just nominated pupils) were asked to complete the assessments at each time point so as not to single out any pupils – particularly important in the control year groups, however only the nominated pupils in each year group were included in the analyses.

Primary outcome

The primary outcome in the trial was the age equivalent score on the Reading module of the InCAS assessment. This module included subsections on decoding, word recognition and comprehension.

Secondary outcomes

Secondary outcomes were age equivalent scores on 1) General Maths and 2) Mental Arithmetic modules as well as mean scores on the attitudes to 3) school, 4) reading and 5) maths subscales measured by the InCAS attitudes module. Total score on the music survey 6) was also a secondary outcome in the second and third years. Total music score per pupil was defined as the sum of the Likert scale (1-4) scores for all the items in the questionnaire. The impact of the programme on these outcomes was estimated to support the analysis of the primary outcome.

Additional exploratory outcomes

Exploratory outcomes were also average point scores at Key Stage 2 (for pupils who had reached this stage) and school absence rate during the year the pupil was involved in the trial (for pupils in trial first and second year due to third year data not being available at the time of writing the report).

Participant selection

School eligibility

Schools were eligible to take part if they were willing to undertake all related research procedures and if they were a 'hub' project school awarded funding by *SHINE* to run the *Hallé SHINE on Manchester* programme, or if they were a local partner school of a school awarded funding. The 'hub' project schools were primary schools (except in one case where a project was funded in a secondary school for pupils in two partner feeder primary schools) recruited by *SHINE* and the Hallé based on a competitive application process open to state schools across Manchester. Applications were assessed under the headings of: 'need and demand' – the nature and level of disadvantage and interest across the partner schools (hub schools needed to have at least 60 eligible children in Year 5 and Year 6 across their cluster, including partner schools); 'capacity to deliver' – ability and capacity to run the project (e.g., 'health' of school, experience, plans for recruitment, strength of partnerships between schools) and 'monitoring and evaluation' – the clarity of outcomes and plans and systems in place to monitor outcomes. Partner schools were approached by the hub schools to take part and were required to be primary schools within walking distance of the hub school which had links to the host school. Head teachers in all schools (hub and partner) gave written informed consent for their school to be involved at the recruitment stage before pre-testing and randomisation (information sheets and consent forms are included in Appendix A).

School recruitment

First Year

The initial recruitment process for pilot schools was completed by *SHINE* and the Hallé working in conjunction with music services in Manchester. Hallé staff initially met with the Head of Music Services for Greater Manchester and then presented the programme at a meeting of Heads of Music from each Manchester borough. Three boroughs were then targeted and *SHINE*'s requirements for schools explained to the Heads of Music for these boroughs. The music services contacted schools and gathered interest. Seven schools expressed an interest at the initial stages and were responded to by phone or email. Four schools continued their interest after receiving further information and were visited by *SHINE*. Three schools then put in a grant application and decisions on whether the schools received funding was made by the *SHINE* board. They assessed all applications as described in the school eligibility section above.

The timeframe for recruitment of schools for the first year of the trial was shorter than the normal grant awarding process due to waiting for confirmation of EEF co-funding and to evaluation planning. *SHINE* made first contact with schools early June 2012 and visited schools late June 2012; schools submitted an application by 18th July 2012; and grants to schools were made on 25th July 2012.

A recruitment meeting was held by *SHINE*, the Hallé and the evaluators on the 10th September 2012 which was attended by staff from all host schools and some of the partner schools. At this meeting, the requirements of the evaluation were set out and details of consent, nominating pupils and testing were explained.

All three host/hub schools had initially applied for the grant with partner schools agreed. For two of the hubs, the partner schools that had been included in the grant application to *SHINE*, consented to and joined the evaluation (one had five partner schools and the other had three partner schools). The third host/hub school had originally applied with agreement from seven partner schools. Four of these partner schools subsequently dropped out of the project before randomisation, two due to not being able to nominate eligible pupils and two due to not being able to fulfil evaluation requirements. Two replacement partner schools were recruited by the host school; however, again both withdrew before randomisation (one due to only being able to nominate pupils from one year group and one to not being able to meet the deadlines of the project). This hub therefore only worked with three partner schools.

Second Year

Recruitment of new schools to the main trial

As in the first year, the recruitment of new schools was completed by *SHINE* and the Hallé in conjunction with music services in Manchester boroughs. Heads of Music services recommended schools that were known to them and these schools were contacted by email by *SHINE*. The aim was to have a *SHINE* project in each Manchester borough. *SHINE* narrowed down the list looking for schools with high FSM and good/outstanding Ofsted ratings. 12 schools expressed an interest in running the intervention and *SHINE* made visits to each interested school. Eight schools then submitted a full application. Grants were made to five hub schools on the same criteria as in the first year. These schools had partner schools already signed up.

The timeframe for recruitment in the second year was longer than in the first year and more representative of the normal grant awarding process. The recruitment period started in February 2013 with the deadline for full applications being at the end of May 2013. Grants were awarded to five successful hub schools on the 13th June 2013 and all schools (hub and partner) were invited to attend an initial meeting in July 2013 with *SHINE*, Hallé and the evaluation team to explain the intervention and the requirements of the evaluation.

Recruitment of additional schools to pilot hubs

Each of the three pilot hubs from the first year recruited additional schools to take part in the second year of the trial (two additional schools in one hub, and one additional school to each of the other hubs). These schools were recruited to allow hubs a greater pool of potential students to attend the intervention with the aim of improving attendance at the intervention. These schools were approached by the project manager or head teacher in the hub school in April or May 2013 to explain the intervention and evaluation. The Hallé *SHINE* programme manager then visited with the head of the new schools to make sure they understood the programme and requirements of the intervention. These schools were invited along to the

initial meeting for new school in July 2013 with the evaluator and consented in the same way as other schools.

Third Year

For the third year, 19 of the 20 main trial schools from the second year continued to deliver the intervention while one school withdrew – this school had stopped engaging with the project by the end of the second year and no children from the school were attending the intervention at the end of the second year. No new schools were recruited.

Pupil eligibility

For pupils to be eligible for the trial they had to attend a participating hub or partner school, be in the appropriate year groups depending on the stage of the trial (see Table 2) and had to have been nominated as eligible for the intervention by teachers before pre-testing and randomisation. Teachers used guidance from the *HSoM* 'Making it Work' handbook to assess eligibility for the programme. Nominated pupils were required to meet at least one of the selection criteria:

- disadvantage as indicated by receipt of free school meals (FSM);
- disadvantage as indicated by another indicator of economic disadvantage;
- children for whom English is an additional language (EAL);
- general underachievement either in relation to their peers or their own potential (but not necessarily low achievement); or
- students with special educational needs (SEN).

Teachers were also asked to only nominate eligible pupils who they thought were likely to be able to attend the intervention.

All pupils in each year group (nominated and not nominated) completed pre-testing. Letters from the evaluator were sent home by schools to inform all parents of the testing, the evaluation and how the data would be used (example at Appendix C). Parents were able to opt out of their child's data being used in the project by contacting the school or research team directly. Schools were asked to send this letter out before randomisation. After randomisation, only the nominated children in the intervention year group were told of the opportunity to attend the *Hallé SHINE on Manchester* programme, and parental consent to attend the intervention was obtained by schools.

Sample size

The number of schools that we expected to work with was pre-specified and restricted by the number of schools funded to host the intervention over the course of the three year evaluation. In the pilot year, three host schools were funded and it was expected that each would have at least three partner schools; therefore, we assumed 12 schools for the first year. These schools were also to be part of the evaluation in 2013/14 while an additional five host schools were also funded with the assumption that each would have at least three partner schools, bringing the number of schools assumed for the main trial up to 32. For 2014/15 the pilot schools were no longer involved and we assumed that 20 schools would remain in the trial.

A sample size calculation was undertaken at the start of the trial for all years of the trial (pilot and main) which made the following assumptions. First, we assumed there would be approximately ten children per year group in each school taking part in the trial. Second, using data from a previous trial undertaken by the lead investigator – the *Every Child Counts* evaluation (Torgerson et al., 2011) – we estimated an intra-cluster correlation coefficient of 0.19 (which was based on a numeracy outcome) and used the pre- post-test correlation data reported for the Progress in Maths 6 which was 0.74. Using the formula $1 + ((\text{class size} - 1) \times \text{intra-cluster correlation})$ gives us a design effect of 2.71.

We estimated that we would have, respectively, at least 120, 320 and 200 pupils in each year group in the pilot, main trial (2013/14) and main trial (2014/15), giving 640 participants in total. However, we had a high pre- post-test correlation and also significant clustering. The pre- post-test correlation inflated the effective sample size to: 260 (i.e., 120 inflated), 694 (i.e., 320 inflated) and 434 (i.e., 200 inflated) giving 1388 in total. However, these then needed to be adjusted downwards by dividing them by the design effect of 2.71. Consequently we achieved actual sample sizes, adjusting for both clustering and pre- post-test correlation, of: 96; 266, 160 for each study respectively and 522 in total. This gave us 80% power to detect differences of: 0.58; 0.35 and 0.45 respectively with a difference of 0.25 when all the groups were pooled.

Randomisation

Details of the recruited schools were provided by *SHINE* to the project team and random assignment was conducted by the project statistician using coded school names so that he was not able to identify the school.

First Year

Randomisation was stratified to take account of the different school types (hub or partner school) and to balance the numbers in the intervention and control groups so that half the schools were allocated to year 5 intervention and half the schools to year 6 intervention both within each hub and overall.

The 14 schools recruited for the pilot trial were stratified by school type into three hub schools, five hub 1 partner schools, three hub 2 partner schools and three hub 3 partner schools. Since 'school type' was the only important factor to be accounted for in the randomisation, the minimisation scheme reduced to randomising schools within stratum to either Year 5 or Year 6 intervention group to ensure equal representation of each stratum in the two arms. To randomly allocate schools within stratum to either Year 5 or Year 6 interventions, permuted block randomisation with a fixed allocation ratio was used. The hub schools were randomly allocated using a permuted block size of three with an allocation ratio of 2:1 in favour of the Year 5 intervention group. The partner schools of hub 1 were allocated to either Year 5 or Year 6 intervention groups using a permuted block size of five with allocation ratio of 2:3. The partner schools of hub 2 were randomly allocated to either Year 5 or Year 6 intervention groups using a block size of 3 with allocation ratio of 1:2 in favour of Year 6 intervention group. The partner schools of hub 3 were randomly allocated to either Year 5 or Year 6 intervention groups using a block size of 3 with an allocation ratio of 2:1 in favour of Year 5 intervention group. Each block size was randomly selected from its entire possible realisation based on the permutation of Year 5 or Year 6 intervention groups according to the specified allocation ratios. The randomisation scheme was implemented in R statistical software.

Second Year

Randomisation in the second year followed a similar process to the first year, but also stratified schools by school level Key Stage 2 obtained from the Ofsted Schools Data Dashboard (Ofsted, c2012-2016) based on feedback from the pilot year from the project statistician.

The 18 pilot schools were stratified by school type into three hub schools, six hub 1 partner schools, four hub 2 partner schools and five hub 3 partner schools. Schools with each hub were matched (ranked) by percentage of pupils achieving level 4 or above in both English and mathematics for the year before the evaluation. To randomly allocate the pilot schools in each stratum to either Year 4 or Year 5 intervention group, permuted block randomisation with a varying block sizes was used to obtain equal number of schools in either group. A similar approach based on permuted block randomisation was used to allocate the 20 main trial schools to either Year 5 or Year 6 intervention group. Based on the allocation scheme, there were ten schools in the Year 5 intervention group and ten schools in the Year 6 intervention group. The randomisation scheme was implemented in R statistical software.

Third Year

Randomisation in the third year followed a similar process to the first year and second years, with schools stratified by hub and partner and by school level Key Stage 2 achievement in the year before the evaluation. A permuted block randomisation scheme was used to allocate the 19 schools to either Year 4 or Year 5 intervention groups. Based on the allocation scheme, there were ten schools in the year 4 intervention group which contained two hub schools and eight partner schools. There were nine schools in Year 5 intervention group which contained two hub school and seven partner schools. The randomisation scheme was implemented in R statistical software.

Analysis

All analyses were conducted on an intention to treat basis, including all randomised children in the groups to which they were originally randomised. Analyses were conducted in SAS version 9.3, using 2-sided significance tests at the 5% significance level.

The statistical analysis plan is included as Appendix E.

Baseline characteristics

Baseline characteristics (gender, free school meal status, baseline reading, baseline general mathematics, baseline mental arithmetic and baseline attitudes scores) were summarised by intervention group across all the schools. Continuous variables (literacy, mathematics and attitudes scores) were summarised using descriptive statistics (n, mean/median). Categorical variables were summarised using frequency counts and percentages by intervention group for each school and across the schools.

Primary outcome

The primary outcome (reading score) was analysed by year group and meta-analysed within each trial to test the null hypothesis that there is no difference in the average reading score between those randomised to *HSoM* and those not randomised to *HSoM* after accounting for baseline reading score, FSM, age, school type and potential clustering within schools.

In the protocol, a generalised estimating equation with exchangeable correlation was proposed for the analysis of the pilot data. However, the analysis method was changed to multi-level modelling to accommodate the EEF policy and requirements for effect size calculation and estimation of intra-cluster correlation for clustering of pupils within schools. The clustering was accounted for by specifying schools as random effects in the model. The model that was fitted for each outcome included the pre-test score as a predictor in each model. Other factors in the models were age, FSM, intervention groups. The datasets for each year group were analysed separately for each year of the trial, and a combined effect of the interventions on all year groups within each trial was obtained by pooling the estimate and effect size for all year groups using a fixed effect meta-analysis model. The effect size for each year group was calculated as the ratio of intervention effects and within-schools variability (SD defined as the square root of error variance; Tymms and Merrell (2003) and Hedges (2007)). The pooled effect size with its associated 95% confidence intervals and pooled estimates together with its 95% confidence intervals are reported for both the primary and the secondary outcomes.

Secondary outcomes

The secondary InCAS outcomes – General Maths, Mental Arithmetic scores and Attitudes – as well as the music survey outcome, were analysed in the same manner as the primary outcome. Each model accounted for age, gender, intervention groups, and school type as well as baseline scores of the corresponding secondary outcome. The pooled effect sizes and average intervention effects are reported for the secondary outcomes.

Exploratory outcomes

Additional analyses exploring the effect of the *HSOM* intervention on KS2 results were also based on multilevel models with school as the clustering variable. Pre-test scores and FSM status were also included in the models. The analyses were performed by trial and by year group with the pooled estimate obtained by fixed effect meta-analyses. Analyses of absence rate were also based on multilevel model with absence rate as the outcome variable. A subgroup analysis was also performed for children eligible for free school meals.

Complier Average Causal Effect (CACE) analyses were based on Schochet and Chiang's (2011) approach, where CACE effect size was estimated through weighting of ITT (intention-to-treat) effect size by the proportion of compliers. The proportion of compliers was calculated as the difference between average compliance in the intervention and control schools based on the proportion of pupils that attended more than a fixed number of *HSOM* sessions. The distribution of complier effect size was generated by performing CACE analyses for a grid of compliance thresholds ranging from at least one session to attending more than 90% of the sessions. A non-parametric bootstrap approach was also carried out to further explore the distribution of complier effect size for year group size where there was a positive effect of the intervention. One thousand bootstrap data were generated by bootstrapping at school level since school was the unit of randomisation.

Implementation and process evaluation methods

Monitoring Implementation

Basic data on the number and timing of intervention sessions offered by each project were submitted to the evaluation team by *SHINE* for each year of the project. We report a summary of this in the process evaluation section of the report.

SHINE also collected pupil attendance records from each project detailing the number of *HSoM* sessions each pupil had attended. These data were shared with the evaluation team in October of each year. We report a summary of attendance at the intervention for each year of the project in the process evaluation section.

Fidelity to the intervention was explored during visits and interviews during the process evaluation.

Process evaluation

Piloting data were collected during the first year of the trial to inform the rest of the trial. The main process evaluation of the *HSoM* programme took place in main trial schools during the second year of the trial (2013/14) and continued during the third year (2014/15).

First Year

Piloting

During the pilot year informal feedback was collected regarding the research design and the intervention. This included feedback from *SHINE*, the Hallé and schools with regard to the evaluation design and processes and any issues with how the programme was running. This was collected through emails and phone calls with *SHINE* and schools throughout the year. Face to face meetings with *SHINE*, the *HSoM* Programme Manager and with school staff involved also took place. School meetings took place in May 2013 at each host school with staff from partner schools also invited to attend. This meeting was chaired by *SHINE* (with input from the evaluator) and gave schools the opportunity to feed back on their experiences of the project and evaluation at that time point.

Two visits to observe the intervention were conducted by different members of the evaluation team: one on a 'culmination day' when the children were performing the music project they had been working on for the term to parents and other visitors, and the other to a standard session when the Hallé musicians were attending. Both these visits were set up by the *HSoM* Programme Manager and were to the same project. These visits were undertaken to help understand how the intervention works and to inform the formal process evaluation to take place in the main trial years.

Convenience sampling was used for the process evaluation during the first year.

Second and Third Years

Design

The process evaluation was split over two academic years 2013/14 and 2014/15. The design was cross-sectional in that observations and perceptions of stakeholders were elicited at set times allowing comparisons to be made between participating projects/schools at specific

time points. As the process evaluation continued over two years, it was also possible to identify and consider developments in the delivery of the programme. Only schools in the main trial were included and where relevant, changes from the pilot programme are noted.

Methods

Five site visits were conducted to *HSoM* project schools (Schools 4,5,6,7 and 8) during the 2014 and 2015 spring and summer terms. This covered all projects in the main trial. The analysis is based on data collected through interviews with key stakeholders (i.e. teachers, project manager, Hallé musicians and education director), lesson observations and focus group interview with pupils carried out by the project researcher, as follows:

- School 5: interview with the project manager, observation of a literacy lesson, a music lesson and a short part of a numeracy lesson, focus-group interview with nine pupils.
- School 6: observation of pupils' practice session for the culmination performance which took place at the end of the day, and interview with the project manager.
- School 4: observation of pupils' practice session for the culmination performance, interviews with the project manager and the Hallé education director.
- School 8: observation of a music lesson, an art lesson and the first half of a science lesson, interview with the Hallé musicians, interview with two teachers and one teaching assistant, focus group interview with three pupils.
- School 7: observation of a music lesson and a short part of two other lessons (puppet making and making lighthouses), interview with the project manager, interview with one of the two musicians and focus group interview with six pupils.

In all Saturday schools observed, the pupils attended three sessions as part of the day (pupils were divided into three groups): one session with an English focus; one with a maths/science focus and one with an artistic focus (music, painting, art, drama and dance). On the days when the Hallé musicians were in school, they offered a music workshop to each group of pupils. There were times, however, when artistic activities would be emphasised more, especially during the weeks before the culmination event when pupils needed to have their artwork ready for the final performance (This was evident in our visit in School 7).

Quality assurance: On two occasions the observations were quality assured through joint visits of project researcher and Principle Investigator. They reached high agreement in their observations in both visits.

'Key person' and stakeholder interviews were also carried out during the third year of the project and these included relevant staff from *SHINE* and the Hallé. These included one interview with the *SHINE* senior programme manager (responsible for *HSoM*) and chief executive at *SHINE* in March 2015, one interview with the Hallé's director for education and the *HSoM* programme manager in April 2015 and two telephone interviews with musicians from the Hallé involved in delivery in July 2015.

Intervention implementation

First Year (pilot)

In the first year two projects ran for 24 Saturday sessions and one ran for 23 Saturday sessions; the projects ran for five hour-long sessions each week during the school term, from November 2012 until July 2013 (the programme was originally designed to last for 25 weeks

but, due to delays with assessment and adverse weather conditions, the programme was shortened).

Second Year (pilot & main trial)

In the second year six projects ran for 25 Saturday sessions and two projects ran for 24 Saturday sessions during the year. With the exception of one project, the programme began in November 2013 and ran through until June 2014. In one main trial project the programme started in January 2014 but ran during school holidays so as to still run for 25 Saturday sessions; the delay in getting started was due to delays in recruiting a project manager.

In all pilot projects and in three main trial projects the programme was run in the funded primary 'host' school (as described above). In one main trial project funding was granted to a secondary school which hosted and ran the project for children from two local feeder primary schools. Another main trial project was located in one of the non-funded primary schools in the cluster with staff employed by the funded school. This was due to the location of the school being more central to the schools in the project.

Third Year (main trial)

In the third year, three projects ran for 25 Saturday sessions and two projects ran for 24 Saturday session during the year. The programme ran from November 2014 to July 2015 in all projects. The projects ran in the same school as for the second year of the project with two exceptions. The project which had run in a partner school during the second year moved into the hub school for 2014/15 to make it easier for staff to implement. In another cluster of schools, the funded hub school changed and the project moved to a new hub school (originally a partner school). This was a joint decision between *SHINE* and the original project due to difficulties in managing the project within the school.

Pupil take up and attendance at intervention

The level of attendance at the intervention is shown in table 3 below.

Table 3. Intervention attendance by nominated pupils

	Number of nominated pupils to attend at least one session (% of group)	Mean number of sessions attended (if attended at least 1)
First (pilot) year	108 (58.4%)	14.6
Second year	371 (51%)	14.1
Third year	224 (48%)	15

In the first (pilot) year out of 185 nominated children in the intervention group 108 (58.4%) attended at least one session and of these pupils the mean number of sessions attended was 14.6, (SD 7.2) out of 23 or 24 possible sessions. In the second year, out of 715 nominated children in the intervention group, 371 (51%) attended at least one *HSoM* session. Of these pupils the mean number of sessions attended was 14.1 (SD 8.4) out for 24 or 25 sessions. In the third year, of the 470 nominated children in the intervention group, 224 (48%) attended at least one *HSoM* session. Of those that attended the mean number of sessions attended was 15 (SD 7.9) out of a possible 24 or 25 sessions.

Table 4 below gives an overview of the process evaluation data collection.

Table 4. Process Evaluation Summary

		Location	Interviews	Observations	Conducted by
2013-14	March 2014	Hub visit	Project Manager; pupil focus group	Literacy, music, numeracy sessions	Project lead; Project researcher
	July 2014	Hub visit	Project manager	Pupil culmination practice & event	Project researcher
2014-15	March 2015	<i>SHINE</i>	<i>SHINE</i> senior programme manager; <i>SHINE</i> chief executive	-	CEM project lead; Trial coordinator
	March 2015	Hub visit	Project manager; Hallé Education Director	Pupil culmination practice	Project lead; Project researcher
	April 2015	Hallé	Education Director and Programme Manager	-	CEM project lead; Trial coordinator
	May 2015	Hub visit	Hallé musicians; 2 teachers & 1 teaching assistant; pupil focus group	Music, art & science lesson	Project researcher
	June 2015	Hub visit	Project manager; Hallé musician; pupil focus group	Music and two other lessons	Project researcher
	July 2015	Telephone interviews	Musicians	-	CEM project lead

Timeline

Table 5 below describes the timeline for the pilot and the main trial RCTs, as well as the process evaluation.

Table 5. Timeline

Date	Activity	
	Pilot Schools	Main trial Schools
June – July 2012	<i>SHINE</i> advertise and invite schools to apply for grants.	
July 2012	<i>SHINE</i> awarded grants to three host schools to run the intervention	
July – October 2012	All schools invited to attend information event and consented to the evaluation	
October 2012	Schools nominated year 5 and year 6 pupils who would be eligible for the intervention and sent the lists to Durham	
October – November 2012	InCAS pre-testing completed and schools told of random assignment	
November 2012 – July 2013	Hallé <i>SHINE</i> on Manchester programme took place each Saturday during term time	
February 2013 – June 2013		New projects hubs and schools recruited to the research by <i>SHINE</i>
June – July 2013	InCAS post-testing completed	<i>SHINE</i> awarded grants to five host schools to run the intervention
July 2013		All new schools invited to attend information event and consented to the evaluation
July 2013/September 2013	Schools sent nominated pupil lists for the year 4 and year 5 groups to <i>SHINE</i> who sent these on to Durham	Schools sent nominated pupil lists for the year 5 and year 6 groups to <i>SHINE</i> who sent these on to Durham
September-October 2013	InCAS pre-testing completed and schools told of random assignment	InCAS pre-testing completed and schools told of random assignment
November – July 2013	Hallé <i>SHINE</i> on Manchester programme took place each Saturday during term time.	Hallé <i>SHINE</i> on Manchester programme took place each Saturday during term time.
June – July 2014	InCAS post-testing completed	InCAS post-testing completed
July – September 2015		Schools sent nominated pupil lists for the year 4 and year 5 groups to <i>SHINE</i> who sent these on to Durham
September – October 2014		InCAS pre-testing completed and schools told of random assignment

November 2014 – July 2015		Hallé <i>SHINE</i> on Manchester programme took place each Saturday during term time.
June – July 2015		InCAS post-testing completed

Costs

Information about the grant awarded to project hub schools was provided by *SHINE*. As this grant was expected to cover all costs to schools associated with delivering the intervention including resources, lunch, school visits as well as teachers' salary costs further cost information was not collected from schools. The budget for schools spending was set out in the *HSoM* 'Making it Work' handbook (*SHINE*, 2012) and schools reported to *SHINE* on any variations in this. Variations to the budget were explored in the process evaluation interview with *SHINE* staff.

The cost per pupil per year was calculated two different ways:

- The second method was to divide the grant by an average of the number of pupils who actually attended the intervention. This is based on the average number of pupils that attended the programme (whether in the trial or not) each week during the third year of the trial when projects and schools were established in delivering the intervention.
- The third method was the expected cost per pupil based on the number of pupils the intervention was designed to accommodate (60).

The evaluation was not able to investigate the additional organisational costs of setting up and running the projects from the *SHINE* end. In the current model of *SHINE* funding schools would not be asked to pay for these. Costs not included in the evaluation:

- Salary/time costs for *SHINE* staff set-up and monitoring
- Salary/time for *HSoM* project manager based at Hallé
- Salary/time for Hallé director of education for music project development
- Salary/time for Hallé musicians to support project
- School maintenance costs

Impact evaluation

Participants

First Year (Pilot)

School and pupil participant flow for the first year (pilot), based on primary outcome data, is shown in Figure 2. In total 370 pupils were nominated at the start of the trial. In the first year of the trial (pilot year) no schools dropped out after randomisation. However, parents of six nominated pupils opted out of their child's data being used in the evaluation and three nominated pupils were not assessed in either the pre- or post-test assessment due to unknown reasons. Forty-seven pupils were missing post-test scores in the primary outcome of reading: five pupils had left the school by the time of post-testing while one pupil was off school until the end of term. Schools did not report on why these remaining pupils did not complete the assessments. This left 314 pupils included in the analyses for primary outcome data at the end of the year. The pupil attrition rate was 15%.

Second Year

School and participant flow for the second year is shown in Figure 3. In total there were 1376 pupils nominated at the start of the trial. One school dropped out of the project during the year after randomisation (30 pupils). The school had very low uptake of the intervention from the nominated pupils and were unwilling to complete the testing at the end of the year. We were also unable to gather follow-up data from another school where the school were unable to coordinate the assessments during the assessment period (23 pupils). An additional school was unable to test the pupils in one year group due to issues with organising the testing (9 pupils). In total 301 pupils were missing post-test scores in the primary outcome assessment (including the 61 described). This left 1075 pupils included in the analysis for primary outcome data at the end of the year. The pupil attrition rate was 22%.

Third Year

School and participant flow for the third year is shown in Figure 4. For the third year, 19 of the 20 main trial schools from the second year continued. There were 930 pupils nominated at the start of the trial. No schools dropped out during the trial. No pupil opt-outs were received. 229 pupils were missing post-test scores in the primary outcome assessment. 701 pupils were included in the analyses for the primary outcome. The pupil attrition rate was 25%.

The level of missing data is higher in the second and third years due to a number of factors. In the pilot year, schools assessed pupils immediately after nominating them as eligible. In the second and third years schools nominated pupils for the intervention before the summer holidays and assessment took place in September and October; therefore there is a higher likelihood that some children would have left the school between these two time points. The assessment software had also changed format between the first and second year of the trial. This meant that schools could use a web-based version of the software which had occasional connection problems showing pupils assessments as incomplete when the internet connection dropped.

First Year (Pilot) Flow Diagram 2012/13

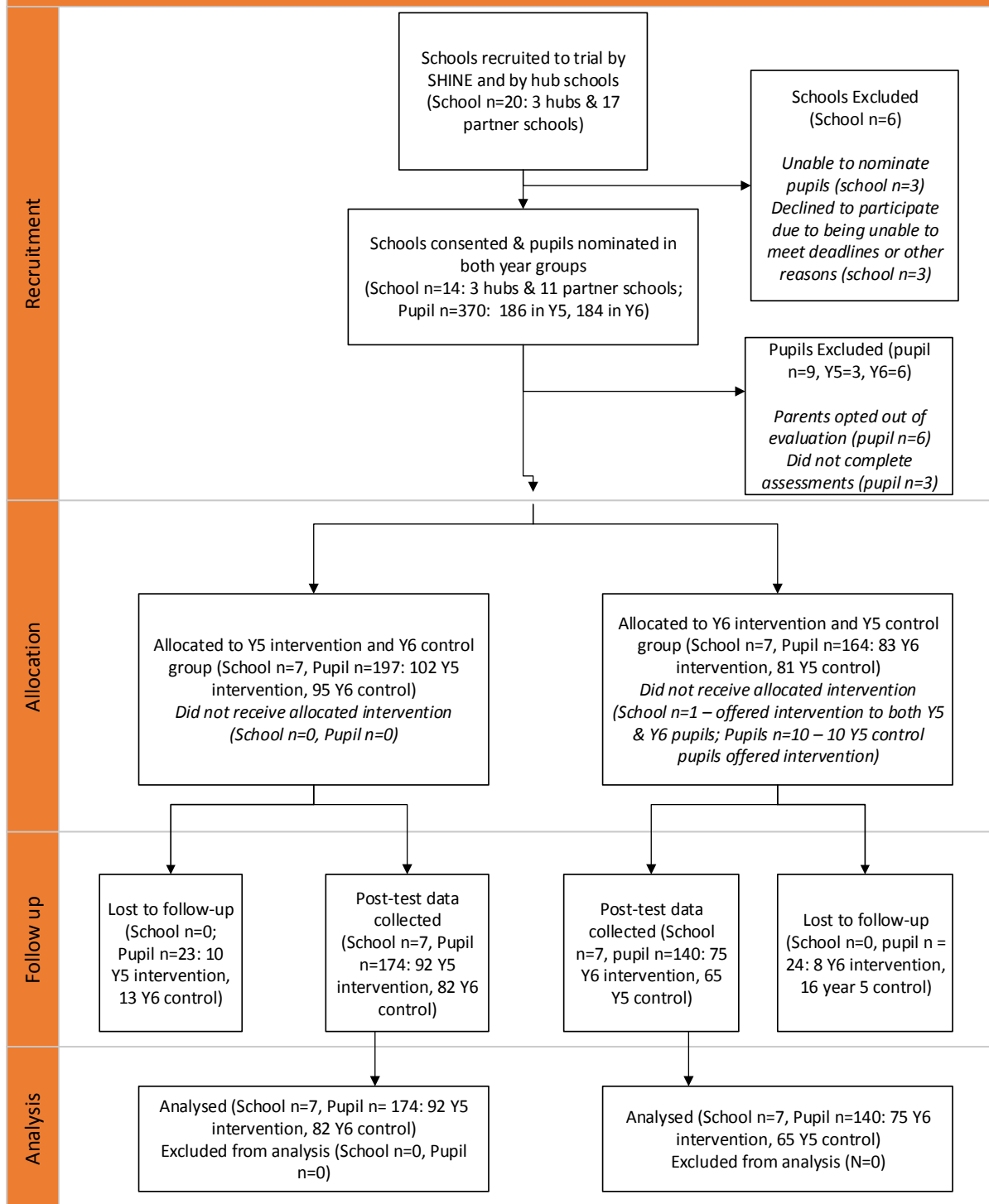


Figure 2. First Year (Pilot Trial) participant flow diagram

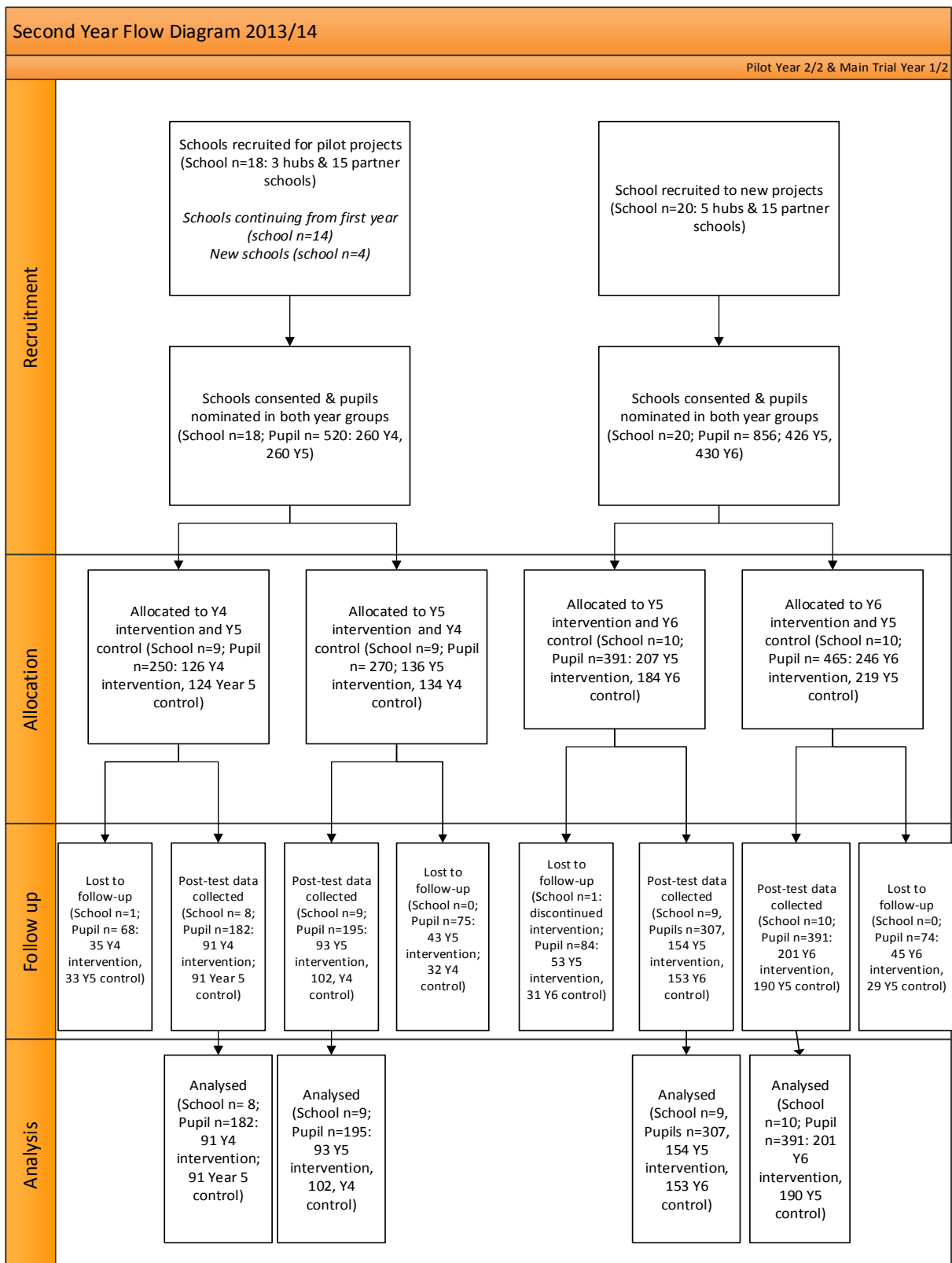


Figure 3. Second Year Participant Flow Diagram

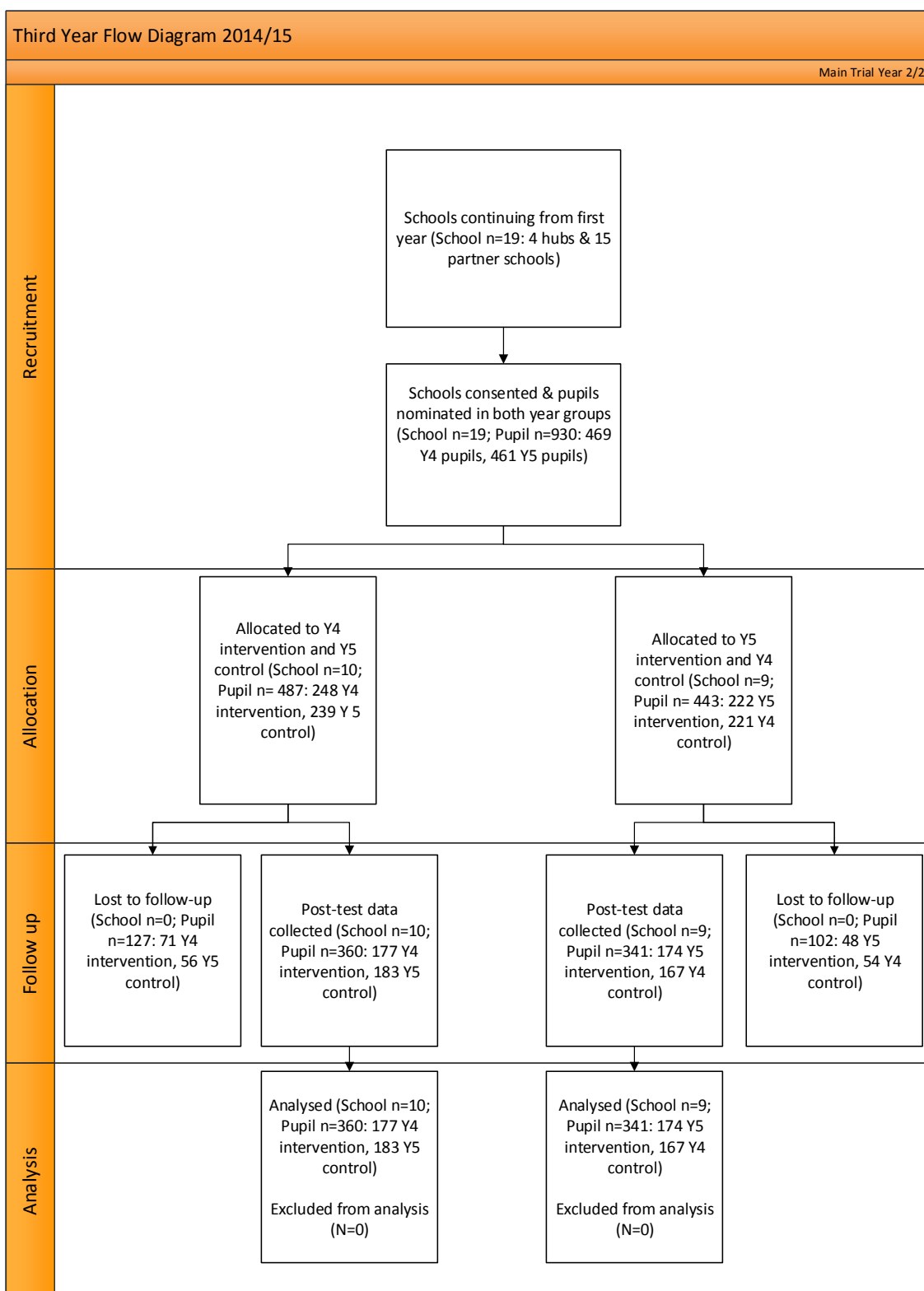


Figure 4. Trial Third Year Participant Flow Diagram

School and Pupil characteristics

First Year

School level baseline characteristics

14 schools participated in the pilot trial with seven schools in the Year 5 intervention group and seven schools in the Year 6 intervention groups. All schools involved were urban schools in the Manchester area; six were local community schools while five were voluntary aided Church of England schools and three were voluntary aided Roman Catholic schools. In their latest Ofsted inspections all three hub schools and seven partner schools were category 2 ('good') while four of the partner schools were category 3 ('requires improvement'). Table 6 below shows school level characteristics for the schools involved.

Table 6. First Year (Pilot) School Level Baseline Characteristics (data from 2012).

		KS2 English L4 + %	KS2 Maths L4+ %	%FSM	%Girls	%EAL	%SEN	Size of School
All schools (N=14)	Mean (SD)	79.71 (12.02)	78.43 (14.3)	36.95 (13.52)	50.17 (3.59)	21.64 (23.8)	9.1 (4.45)	278.79 (80.93)
	Range	57-98	46-100	16.7- 64.3	42.6- 55.8	2.2-76.8	4.3-18.5	184-470
Hub schools (N=3)	Mean (SD)	78.33 (19.43)	78 (27.73)	32.93 (14.08)	52.97 (2.65)	10.57 (10.61)	8 (0.62)	263.33 (40.22)
	Range	57-95	46-95	16.7- 41.9	50.3- 55.6	2.2-22.5	7.5-8.7	230-308
Partner schools (N=11)	Mean (SD)	80.09 (10.57)	78.55 (10.55)	38.05 (13.85)	49.40 (3.51)	24.66 (25.83)	9.4 (5.02)	283 (89)
	Range	60-98	64-100	19-64.3	42.6- 55.8	5-76.8	4.3-18.5	184-470

Pupil level baseline characteristics

Baseline characteristics for pupils in the pilot trial are presented in Table 7. There were 361 nominated eligible pupils in Year 5 and Year 6 with 185 in the intervention groups and 176 in the control group. There was no gender difference at baseline with 48% and 45% male pupils in the intervention and control groups, respectively. Also, the proportion of children with free school meal (FSM) status, were similar between the intervention and control groups, respectively. At baseline, although the 'attitude outcomes', were not very different between intervention and control groups, the variability within group was quite big compared to other indicators. All other outcomes seemed comparable between the intervention and the control groups. There were 3% and 10% missing cases in the primary outcome for the intervention and the control groups, respectively.

Table 7. Description of pupil level characteristics (first year)

Baseline Comparison				
Variable	Intervention group		Control group	
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for FSM	126/185(0)	68%	116/176(0)	66%
Male	88/185(0)	48%	79/176(0)	45%
Pupil-level (continuous)	n (missing)	Mean±SD	n (missing)	Mean±SD
Reading	185(6)	9.18±1.65	176(17)	9.63± 1.58
Mental Arithmetic	185(6)	9.29±1.54	176 (6)	9.34± 1.54
General Maths	185(8)	9.25± 1.26	176 (6)	9.41± 1.25
Maths	185(4)	44.96±35.27	176 (4)	39.75±40.12
Attitudes to Maths	185(4)	47.29±39.75	176 (4)	57.97±33.52
Attitudes to Reading	185(4)	68.91±27.37	176 (4)	65.85±29.46
Attitudes to School	185(4)			
Sch				

Second Year

School level baseline characteristics

38 schools took part in the second year of the project (14 of the 38 schools had also been involved in the *SHINE* pilot study). All were urban schools in the Greater Manchester area. Four schools were sponsor-led academies, 13 were Church of England schools (seven were voluntary aided and six voluntary controlled schools), 16 were community schools and five were voluntary aided Roman Catholic schools. In their latest Ofsted inspections three partner schools were rated 'outstanding', 27 schools (three hubs and 24 partner schools) were rated 'good' and 8 schools (one hub and seven partner schools) were rated 'requires improvement'. Table 8 below shows school level characteristics of the schools involved in the second year of the trial.

Table 8. Second year school level characteristics (data retrieved from Ofsted data dashboard (Ofsted, c2012-2016) for 2013)

		KS2 Reading L4 + %	KS2 Maths L4+ %	%FSM	%Girls	%SEN	Size of School
All schools (N=38)	Mean (SD)	81.55 (11.25)	83.92 (8.25)	46.24 (14.41)	49.11 (3.04)	7.75 (3.65)	315.53 (104.62)
	Range	54-100	63-100	14-73.2	41.6-55	1.7-16.8	182-584
Hub schools (N=7)	Mean (SD)	78.28 (12.62)	79.57 (9.83)	45.83 (11.78)	49.37 (3.56)	8.19 (3.45)	304 (116.13)
	Range	54-91	63-93	26.2-59.2	43.7-54.4	4-13.2	182-469
Partner schools (N=31)	Mean (SD)	82.29 (11.20)	84.90 (7.85)	46.34 (15.34)	49.05 (3.02)	7.65 (3.80)	318.13 (105.49)
	Range	60-100	64-100	14-73.2	41.6-55	1.7-16.8	189-584

**EAL data were not available at the time of gathering data*

Second year pupil level baseline characteristics

There were 1376 pupils in the second year main trial with 715 and 661 pupils in the intervention and the control groups, respectively. Between the intervention and control groups, the proportion of male pupil (49% and 51%, respectively) and those with free school meal (FSM) status (67% for both) seemed quite similar. Between the intervention and the control groups, all outcomes were quite similar. The attitude related outcomes were also similar between the two groups, but there were more variability within the groups. There were 5% missing cases in the primary outcome for the intervention and the control groups. The summary of the baseline characteristics are presented in Table 9.

Table 9. Description of pupils level characteristics (second year)

Baseline Comparison				
Variable	Intervention group		Control group	
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for FSM	478/715(8)	67%	443/661(1)	67%
Male	351/715(23)	49%	340/661(18)	51%
Pupil-level (continuous)	n (missing)	Mean±SD	n (missing)	Mean±SD
Reading	715(38)	8.96± 1.81	661(32)	9.06± 1.77
Mental Arithmetic Meth	715(35)	8.64±1.85	661(55)	8.60± 1.79
General Maths Maths	715(34)	8.86± 1.26	661(36)	8.80± 1.27
Attitudes to Maths Maths	715(63)	38.35± 43.24	661(74)	39.59± 42.80
Attitudes to Reading Reading	715(63)	43.04± 44.99	661(74)	43.48± 44.81
Attitudes to Schools	715(63)	61.32± 35.99	661(74)	57.19± 37.47

Third Year

Third Year school level baseline characteristics

The third year trial consisted of 19 schools all of which were urban schools in the Greater Manchester area. Two were sponsor-led academies, eight were Church of England schools (four were voluntary aided and four were voluntary controlled schools), seven were local community schools and two were voluntary aided Roman Catholic schools. In their latest Ofsted inspections, one school was rated 'outstanding' (hub), sixteen were rated 'good' (two hubs, fourteen partner schools), one was rated 'requires improvement' (hub), one was rated 'inadequate' and placed into special measures (partner school).

Table 10. Third year school level characteristics (data retrieved from Ofsted data dashboard (Ofsted, c2012-2016) for 2014)

		KS2 Reading L4 + %	KS2 Maths L4+ %	%FSM	%Girls	%SEN	Size of School
All schools (N= 19)	Mean	84.11	84.58	45.42	46.12	7.87	339.79
	(SD)	(6.53)	(7.61)	(12.05)	(10.91)	(3.03)	(122.87)
	Range	67-93	70-94	21.7-68.9	45.5 -52.0	2.3-14.1	192 - 591
Hub (N=4)	Mean	85.25	88.25	47.55	48.03	5.73	408.50
	(SD)	(3.49)	(2.59)	(9.92)	(1.82)	(2.52)	(108.34)
	Range	82 - 91	84 – 91	36.4–59.1	45.5–49.9	2.3–9.4	221 - 478
Partner Schools (n= 15)	Mean	83.8	83.6	44.85	45.61	8.44	321.47
	(SD)	(7.09)	(8.19)	(12.5)	(12.19)	(2.89)	(120)
	Range	67 - 93	70 - 94	21.7–68.9	46.1–52.0	3.4–14.1	192 - 591

**EAL data were not available at the time of gathering data*

Third Year pupil level baseline characteristics

There were 930 pupils in the third year trial with 470 and 460 pupils in the intervention and the control groups, respectively. Similar to previous years, the distribution of gender and free school meal status (FSM) among the intervention and control group were quite similar. Apart from the 'attitude to school, all other outcomes were comparable at baseline between the intervention and the control groups. There were 7% missing cases in the primary outcome in the intervention and the control groups, respectively. The summary of the baseline characteristics are presented in Table 11.

Table 11. Description of pupil level characteristics (third year)

Baseline Comparison				
Variable	Intervention group		Control group	
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for FSM	282/470(4)	60%	285/460(3)	61%
Male	245/470(4)	52%	247/460(1)	54%
Pupil-level (continuous)	n (missing)	Mean±SD	n (missing)	Mean± SD
Reading	470(34)	8.45± 1.85	460(34)	8.54± 1.98
Mental Arithmetic				
Meth	470(51)	8.01± 1.77	460(52)	7.96± 1.94
General Maths				
Maths	470(34)	8.39± 1.21	460(34)	8.43± 1.34
Attitudes to Maths				
Maths	470(56)	44.40± 45.75	460(55)	45.94±41.89
Attitudes to Reading				
Reading	470(56)	50.51± 45.37	460(55)	50.11±39.17
Attitudes to School				
	470(56)	8.45± 2.34	460(55)	8.34± 2.52

Outcomes and analysis

First Year

Primary outcome

As pre-specified in the approved protocol, the pilot data were analysed at the end of the first (pilot) year; similarly, with the second and third years' data. These results were then combined in the meta-analysis as agreed by all parties.

The pilot data were analysed separately for Year 5 and Year 6 pupils using multi-level models; schools were specified as random effects to account for heterogeneity between schools. The effect sizes were pooled into a single value using fixed-effect meta-analysis model. The effect size for each year was calculated as the ratio of the intervention effects and total variance using the approach of Hedges (2007). The pooled effect size of Year 5 and Year 6 intervention effects are reported in Table 12. The pooled effect size was estimated as 0.03(-0.27, 0.34), which was not statistically significant at 5% level. There was no evidence to suggest that the intervention improves pupils' reading. The percentage of post-test missing data (second year) was comparable between the intervention and the control groups. 7% (13) of the 185 pupils in the intervention group and 8% (14) of the 176 pupils in the control group had post-test missing data. The overall percentage of post-test missing data was 7%.

Table 12. Pooled effect size and estimate together with 95% confidence intervals for the primary outcome (reading)

Primary analysis						
Outcome	Raw means				Effect size	
	Intervention group	Control group				
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g** (95% CI)
Post-test	185(13)	9.80(9.56, 10.05)	176(14)	10.08(9.83, 10.32)	314(167, 147)	0.03(-0.27, 0.34)

Secondary outcomes

The secondary outcomes were analysed using the same approach used for the primary outcome. The data were first analysed separately for Year 5 and Year 6, then the effect sizes were combined in a fixed effect meta-analysis model to obtain a single effect for each outcome. The effect sizes were calculated as the standardized difference between the intervention and the control groups using total variance. The results are presented in Table 13. The effect sizes range from -0.24 ('attitudes to reading' scores) to 0.04 ('attitudes to school' scores). None of the secondary outcomes showed a significant positive or negative effect of the intervention.

Table 13. Pooled effect size from intention-to-treat analysis of the secondary outcomes using multi-level model

Secondary analysis							
Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges (95% CI)	g**
Mental Arith	185(8)	9.65(9.43, 9.88)	176(9)	9.76(9.53, 9.98)	33(172, 161)	-0.04(-0.32, 0.24)	
Gen Maths	185(12)	9.77(9.57, 9.98)	176(9)	10.02(9.81, 10.23)	326(165, 161)	-0.15(-0.45, 0.14)	
Att. Maths	185(7)	44.96(39.41, 50.52)	176(12)	39.75(33.96, 45.54)	342(178, 164)	0.07(-0.19, 0.33)	
Att. Reading	185(7)	47.29(41.85, 52.73)	176(12)	57.97(52.31, 63.63)	334(174, 160)	-0.24(-0.46, 0.01)	-
Att. School	185(7)	59.26(54.34, 64.18)	176(12)	59.54(54.41, 64.47)	334(174, 160)	0.04(-0.19, 0.27)	

MAIN Trial

Second Year

Primary outcome

The main trial data were analysed separately for Year 4, Year 5, and Year 6 pupils using multi-level models; schools were specified as random effects to account for heterogeneity between schools. The effect sizes were pooled into a single value using fixed-effect meta-analysis model. The effect size for each class was calculated as the ratio of intervention effects and total variance using the approach of Hedges (2007). The pooled effect size is reported in Table 14. There was no evidence to suggest that the intervention improves pupils' reading. The pooled effect size was estimated as -0.10(-0.26, 0.07). The percentage of post-test missing data was comparable between the intervention and the control groups. 23% (162) of the 715 pupils in the intervention group and 17% (112) of the 661 pupils in the control group had post-test missing data. The overall percentage of post-test missing data was 20%.

Table 14. Pooled effect size and estimate together with 95% confidence intervals for the primary outcome (reading)

Primary analysis							
Outcome	Raw means				Effect size		
	Intervention group		Control group		n in model (intervention; control)	Hedges (95% CI)	g**
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*			
Post-test	715(162)	9.65(9.50, 9.80)	661(112)	9.78(9.63, 9.34)	1071(536, 535)	-0.10(-0.26, 0.07)	

Secondary outcomes

The secondary outcomes were analysed using the same approach used for the primary outcome. The data were first analysed separately for Year 4, Year 5 and Year 6. Then the effect sizes were combined in a fixed effect meta-analysis model to obtain a single effect per outcome. Effect sizes were calculated as the standardized mean difference between the intervention and the control groups using total variance. The results are presented in Table 15. The effect sizes ranged from -0.28 ('general mathematics' scores) to 0.02.('attitudes to reading' scores). None of the secondary outcomes showed a significant effect of the intervention.

Table 15. Pooled effect size from intention-to treat analysis of the secondary outcomes using multi-level model

Secondary Analysis						
Outcome	Raw means				Effect size	
	Intervention group	Control group				
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g** (95% CI)
Mental Arith.	715(134)	9.36(9.21, 9.51)	661(111)	9.39(9.23, 9.54)	1082(566, 516)	-0.10(-0.26, 0.06)
Gen. Maths	715(135)	9.60(9.48, 9.71)	661(110)	9.69(9.57, 9.81)	1102(565, 537)	-0.28(-0.43, -0.13)
Att.Maths	715(128)	38.35(34.86, 41.83)	661(118)	39.59(35.97, 43.22)	1126(584, 542)	-0.04(-0.22, 0.14)
Att.Reading	715(128)	43.04(39.40, 46.67)	661(118)	43.48(39.70, 47.26)	1049(552, 497)	0.02(-0.13, 0.16)
Att.School	715(128)	55.70(52.73, 58.67)	661(118)	55.30(52.21, 58.39)	1049(552, 497)	-0.03(-0.20, 0.15)
Music	715(411)	41.62(40.06, 43.17)	661(370)	42.58(41.00, 44.17)	593(303, 290)	-0.09(-0.37, 0.19)

Minimum detectable effect size

Table 16 presents the estimated minimum detectable effect size for the reading outcome at protocol, randomisation and analysis stages. To be consistent with a single pooled intervention effect per outcome, the MDES was based on the total number of schools and pupils for Year 4, Year 5 and Year 6. The estimated pooled effect size of -0.10 is smaller than the expected MDES of 0.31 at protocol stage, 0.28 at randomisation stage and 0.26 at analysis stage. (Randomisation numbers refer to pupils recruited in that year as per the protocol, the higher analysis numbers reflect the inclusion of pupils from the previous years.)

Table 16. Minimum detectable effect size for second year trial

Minimum detectable effect size at different stages							
Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	640(320,320)	0.74	0.19	School blocking, schools	32 80%	0.05	0.31
Randomisation	860 (430; 430)	0.74	0.19	School blocking, schools	38 80%	0.05	0.28
Analysis (i.e. pre available and post test)	1071(536, 535)	0.69	0.14	School blocking, schools	35 80%	0.05	0.26

Third Year

Primary outcome

The third year trial data were also analysed separately for Year 4 and Year 5 pupils using multi-level models with schools specified as random effects to account for heterogeneity between schools. The pooled effect of the intervention was obtained based on a fixed-effect meta-analysis model. The effect size for each was calculated as the ratio of intervention effects and total variance using the approach of Hedges (2007). The pooled effect size is reported in Table 17. There was no evidence to suggest that the intervention improves the pupils' reading. The pooled effect size was estimated as 0.10(-0.20, 0.40). The percentage of post-test missing data was comparable between the intervention and the control groups. 24% (113) of the 470 pupils in the intervention group and 22% (101) of the 460 pupils in the control group had post-test missing data. The overall percentage of post-test missing data was 23%.

Table 17. Pooled effect size together with 95% confidence intervals for the primary outcome (reading) for year 3 trial.

Primary analysis						
Outcome	Raw means				Effect size	
	Intervention group		Control group		n in model (intervention; control)	Hedges g** (95% CI)
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*		
Post-test	470(113)	9.59(9.41, 9.78)	460(101)	9.59(9.40, 9.77)	699(350,349)	0.10(-0.20, 0.40)

Secondary outcomes

The secondary outcomes were analysed using the same approach as the primary outcome. The data were first analysed separately for Year 4 and Year 5. Then the effect sizes were combined in a fixed effect meta-analysis model to obtain a single effect for each outcome. Effect sizes were calculated as the standardized mean difference between the intervention and the control groups using total variance. The results are presented in Table 18. The effect sizes ranged from -0.17 ('attitudes to school' scores) to -0.01 ('arithmetic' scores). None of the secondary outcomes showed a significant effect of the intervention.

Table 18. Effect size and estimates from intention-to-treat analysis of the secondary outcomes using multi-level model

Secondary Analysis								
Outcome	Raw means				Effect size			
	Intervention group			Control group				
	n (missing)	Mean (95% CI)*		n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g** (95% CI)	
Mental Arith.	470(120)	9.06(8.88, 9.26)		460(106)	8.93(8.74, 9.12)	678(342, 336)	-0.01(-0.26, 0.23)	
Gen. Maths	470(110)	9.15(9.01, 9.30)		460(95)	9.11(8.97, 9.25))	706(351, 355)	-0.02(-0.23, 0.19)	
Att. Maths	470(100)	44.40(39.92, 48.88)		460(96)	45.94(41.41, 50.45)	732(369, 363)	-0.12(-0.30, 0.05)	
Att. Reading	470(100)	50.51(46.19, 54.84)		460(97)	50.11(45.74, 54.48)	729(368, 361)	-0.03(-0.21, 0.16)	
Att. School	470(100)	52.52(48.48, 56.55)		460(97)	58.72(54.64, 62.79)	729(368, 361)	-0.17(-0.41, 0.07)	
Music	470(157)	57.09(55.29, 58.65)		460(164)	57.98(56.38, 59.59)	608(313, 295)-	-0.08(-0.37, 0.21)	

Minimum detectable effect size

Table 19 presents the estimated minimum detectable effect size for the reading outcome at protocol, randomisation and analysis stages. To be consistent with a single pooled intervention effect per outcome, the MDES was based on the total number of schools and pupils for Year 4, Year 5 and Year 6. The estimated pooled effect size of 0.10 is smaller than the expected MDES of 0.40 at protocol stage, 0.41 at randomisation stage and 0.36 at analysis stage. (Randomisation numbers refer to pupils recruited in that year as per the protocol, the higher analysis numbers reflect the inclusion of pupils from the previous years.)

Table 19. Minimum detectable effect size for third year trial

Minimum detectable effect size at different stages								
Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)	
Protocol	400 (200; 200)	0.74	0.19	School blocking, schools	20 . 80%	0.05	0.40	
Randomisation	380 (190; 190)	0.74	0.19	School blocking, schools	19 80%	0.05	0.41	
Analysis available pre and post test)	(i.e. pre 699(349, 350)	0.71	0.16	School blocking, schools	19 80%	0.05	0.36	

Analysis of primary outcome by class

Analysis of primary outcome by year groups is presented in Table 20 and Figure 5. There was no significant effect of the interventions in any of the classes across the different trials. Most of the variability in the pilot trial was within school rather than between schools, proportion of the total variability attributable to between school was 2% for Year 5 (ICC=0.02) and 9% for Year 6 (ICC=0.09). In the second year trial, there was 0% ICC for year 4, 5% for Year 5 and 14% for Year 6 data. In the third year trial, there was 10% ICC for Year 4 and 16% for Year 5 data.

Table 20. Analysis of primary outcome by year group

Primary analysis for Pilot Study							
Year group	Outcome	Raw means				Effect size	
		Intervention group		Control group		n in model (intervention; control)	Hedges g** (95% CI) r.Int
		n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*		
5	Post-test	102(7)	9.51(9.19, 9.83)	81(5)	9.65(9.29, 10.01)	157(92,65)	0.30(-0.11, 0.71)
6	Post-test	83(6)	10.17(9.82, 10.51)	95(9)	10.45(10.12, 10.78)	157(75,82)	-0.29(-0.74, 0.16)
Primary analysis for Second Year Trial							
Year group	Outcome	Raw means				Effect size	
		Intervention group		Control group		n in model (intervention; control)	Hedges g** (95% CI)
		n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*		
4	Post-test	126(33)	8.85(8.52, 9.17)	134(26)	9.32(9.02, 9.63)	192(90, 102)	-0.10(-0.37, 0.17)
5	Post-test	343(88)	9.47(9.25, 9.68)	343(56)	9.85(9.65, 10.06)	526(246, 280)	-0.20(-0.44, 0.04)
6	Post-test	246(41)	10.25(10.00, 10.50)	184(30)	9.98(9.68, 10.27)	353(200, 153)	0.07(-0.36, 0.50)
Primary analysis for Third Year Trial							
Year group	Outcome	Raw means				Effect size	
		Intervention group		Control group		n in model (intervention; control)	Hedges g** (95% CI)
		n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*		
4	Post-test	248(66)	9.32(9.06, 9.58)	221(50)	9.03(8.77, 9.30)	344(177, 167)	0.19(-0.20, 0.58)
5	Post-test	222(47)	9.88(9.62, 10.14)	239(51)	10.09(9.84, 10.34)	355(173, 182)	-0.02(-0.47, 0.43)

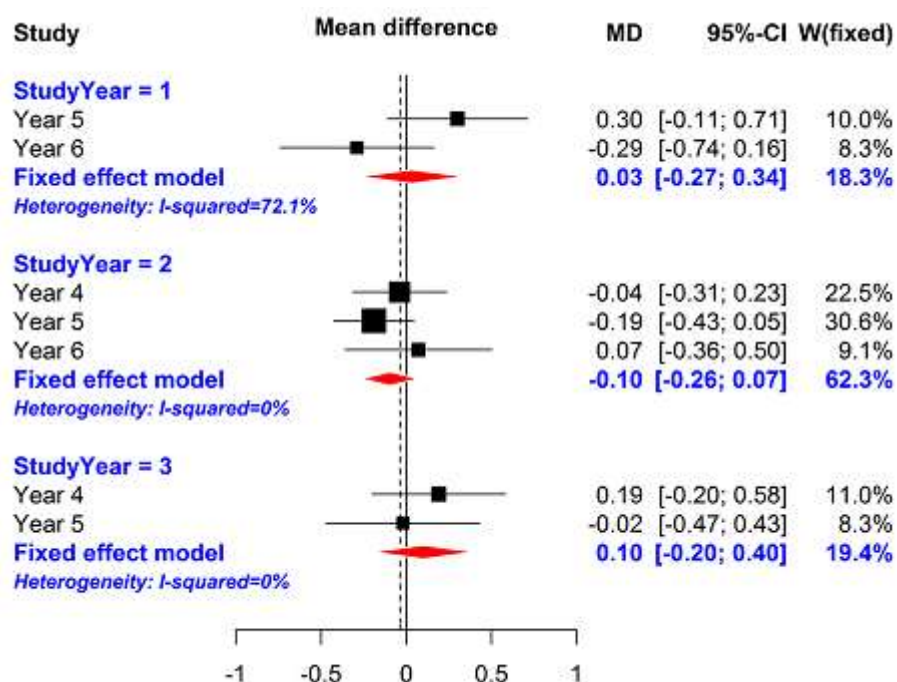


Figure 5. Analysis of primary outcome by year group

Subgroup analysis of primary outcome for FSM

Subgroup analyses of primary outcome for pupils on FSM are presented in Table 21. There was no significant effect of the interventions in any of the trials. The effect sizes for FSM subgroup ranges from -0.13 for the second year trial to 0.05 for the pilot trial.

Table 21. Subgroup analysis of primary outcome for FSM pupils

FSM Subgroup Analysis							
Year 1 Pilot Study							
Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n	Mean	n	Mean	n in model	Hedges	g**
	(missing)	(95% CI)*	(missing)	(95% CI)*	(intervention; control)	(95% CI)	
Post-test	126(12)	9.74(9.44, 10.04)	116(8)	10.10(9.79, 10.41)	212(111, 101)	0.05(-0.27, 0.38)	
Year 2 Main Trial							
Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n	Mean	n	Mean	n in model	Hedges	g**
	(missing)	(95% CI)*	(missing)	(95% CI)*	(intervention; control)	(95% CI)	
Post-test	478(114)	9.56(9.37, 9.74)	443(83)	9.66(9.48, 9.84)	700(350, 350)	-0.13(-0.33, 0.07)	
Year 3 Main Trial							
Outcome	Raw means				Effect size		
	Intervention group		Control group				
	n	Mean	n	Mean	n in model	Hedges	g**
	(missing)	(95% CI)*	(missing)	(95% CI)*	(intervention; control)	(95% CI)	
Post-test	282(74)	9.62(9.39, 9.86))	285(68))	9.48(9.25, 9.71)	413(203, 210)	0.01(-0.29, 0.31)	

Change to analysis plan

The analysis method was changed from generalised estimating equations to multi-level model to accommodate new EEF policy and requirements. Although generalised estimating equations (GEE) corrects for clustering, it does not explicitly estimate the dependency between pupils from the same school. Also, the estimation of effect size is problematic because GEE does not decompose variability into different sources. The predictors in the model were also changed to include FSM. Age at pre-test was not included in the analysis because of limited variability in the data. Gender was not included to accommodate FSM and prevent excluding cases with missing gender data, particularly since including gender made very little difference to the estimated intervention effects. The data were analysed in R using lme4, meta and metaphor packages instead of SAS as stated in the analysis protocol.

Additional exploratory analysis

Additional analyses were performed to explore the effect of the intervention on Key Stage 2 (KS2) results for pupils in year 5 and year 6 in the first two years of the trials. The pooled effect size of the *SHINE* intervention on KS2 reading scores was -0.04(-0.44, 0.35) for the pupils in the pilot trial and 0.00(-0.21, 0.22) for the pupils in the second year trial. The estimated effect for KS2 maths was -0.08(-0.43, 0.28) during pilot trial and -0.05(-0.23, 0.14) during the second year trial. The estimated effect for KS2 writing was -0.27(-0.56, 0.02) during the pilot trial and -0.09(-0.33, 0.14) during the second year trial. The data are consistent with an overall no difference in effect for KS2 reading, maths and writing.

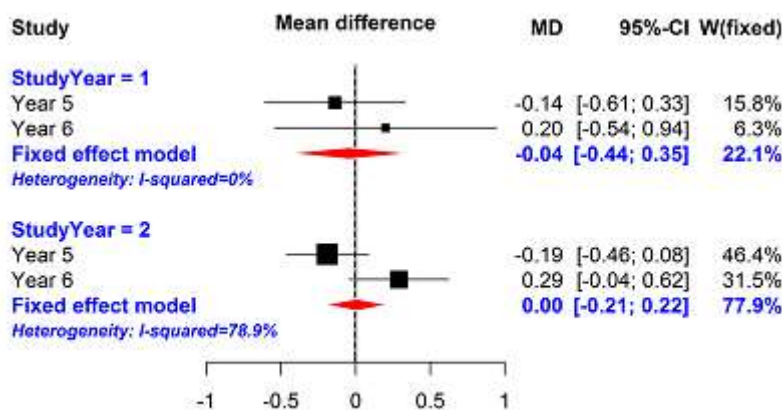


Figure 6. Exploratory analysis of intervention effect on KS2 reading

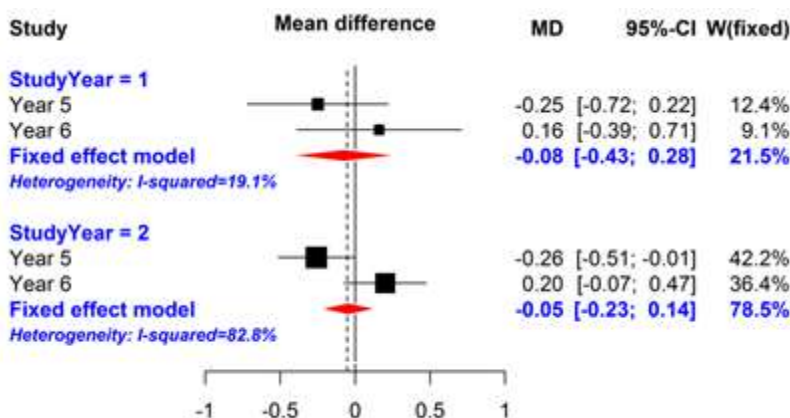


Figure 7. Exploratory analysis of intervention effect on KS2 maths

Additional analyses of the intervention on school absence rate were performed for the pupils in the second year trial by year group. The mean (SD) absence rate for Year 4 were 3.40(3.47) and 4.80(4.75) for the intervention and control groups, respectively. For Year 5, the mean (SD) absence rate were 4.15(4.05) and 3.67 (4.11) for the intervention and control groups, respectively. The mean (SD) absence rate for year 6 were 4.19(4.47) and 3.24(3.24) for the intervention and control groups, respectively. The estimated difference between the intervention and control groups were -1.63(-3.54, 0.20), 0.64(-0.30, 1.59) and 0.87(-0.15,

1.89) for Year 4, 5 and 6, respectively. These data show no statistically significant difference in school absence rate for any of the years.

Exploratory analysis: CACE analysis for primary outcome

Table 22 describes the distribution of average compliance in the intervention and control schools when average attendance was more than a fixed number of SHINE sessions. During the pilot trial, an average of 63% of year 5 pupils in the intervention schools and 9% of year 5 pupils in the control schools attended at least one SHINE session. Among the year 6 pupils in the intervention schools, on average 52% attended at least one SHINE session. In the second year trial, percentages of pupil who attended at least one session in the intervention schools were 62%, 58% and 48% among Year 4, 5 and 6 pupils, respectively. In the third year trial, 48% of Year 4 and 51% of Year 5 pupils in the intervention groups attended at least one SHINE session. In general, the uptake of the intervention was low and only about 50% of the pupils in the intervention school attended at least one *SHINE* session.

Table 22. Average compliance in intervention and control schools, complier (ϕ) was the difference in average compliance between the intervention and the control groups.

Group	Compliance Thresholds (%)									
	p>0	p>10	p>20	p>30	p>40	p>50	p>60	p>70	p>80	p>90
Pilot Trial Year Group 5										
Intervention(T)	0.63	0.62	0.55	0.49	0.47	0.43	0.42	0.41	0.32	0.16
Control(C)	0.09	0.07	0.07	0.07	0.07	0.07	0.07	0.05	0.04	0.02
$\phi = T-C$	0.54	0.54	0.48	0.42	0.4	0.36	0.35	0.35	0.28	0.14
Pilot Trial Year Group 6										
Intervention(T)	0.52	0.49	0.44	0.41	0.38	0.32	0.31	0.26	0.2	0.14
Control(C)	0	0	0	0	0	0	0	0	0	0
$\phi = T-C$	0.52	0.49	0.44	0.41	0.38	0.32	0.31	0.26	0.2	0.14
Second Trial Year Group 4										
Intervention(T)	0.62	0.55	0.46	0.38	0.33	0.30	0.24	0.19	0.19	0.13
Control(C)	0	0	0	0	0	0	0	0	0	0
$\phi = T-C$	0.62	0.55	0.46	0.38	0.33	0.30	0.24	0.19	0.19	0.13
Second Trial Year Group 5										
Intervention(T)	0.58	0.52	0.46	0.44	0.41	0.37	0.36	0.33	0.27	0.17
Control(C)	0	0	0	0	0	0	0	0	0	0
$\phi = T-C$	0.57	0.52	0.46	0.44	0.4	0.37	0.35	0.32	0.27	0.17
Second Trial Year Group 6										
Intervention(T)	0.48	0.45	0.37	0.35	0.31	0.27	0.26	0.24	0.18	0.11
Control(C)	0	0	0	0	0	0	0	0	0	0
$\phi = T-C$	0.48	0.45	0.37	0.35	0.31	0.27	0.26	0.24	0.18	0.11
Third Trial Year Group 4										
Intervention(T)	0.51	0.48	0.41	0.39	0.37	0.34	0.3	0.27	0.15	0.12
Control(C)	0	0	0	0	0	0	0	0	0	0
$\phi = T-C$	0.51	0.48	0.41	0.39	0.37	0.34	0.3	0.27	0.15	0.12
Third Trial Year Group 5										
Intervention(T)	0.46	0.4	0.38	0.35	0.34	0.30	0.26	0.23	0.19	0.12
Control(C)	0.03	0.01	0	0	0	0	0	0	0	0
$\phi = T-C$	0.43	0.39	0.38	0.35	0.34	0.30	0.26	0.23	0.19	0.12

Figure 8 shows the Complier Average Causal Effect (CACE) analysis among those who attended more than 50% of sessions (compliance threshold). Complier Average Causal Effect (CACE) analysis was used to ascertain any relationship between outcome and adherence to the intervention. Estimates of reading attainment for pupils who adhered to the *SHINE* intervention were found to be different from the primary estimates, demonstrating greater benefit amongst adherers in the pilot phase and in the third year and greater detriment amongst adherers in the second year, although none of the results was statistically significant. In the CACE analysis, the pooled effect size of the *SHINE* intervention on reading attainment was 0.06 (95% CI: -7.22 to 7.35) for the pilot phase, -0.39 (95% CI: -5.90 to 5.12) for the pupils in the second year trial and 0.29 (95% CI: -7.76 to 8.34) for pupils in the third year trial.

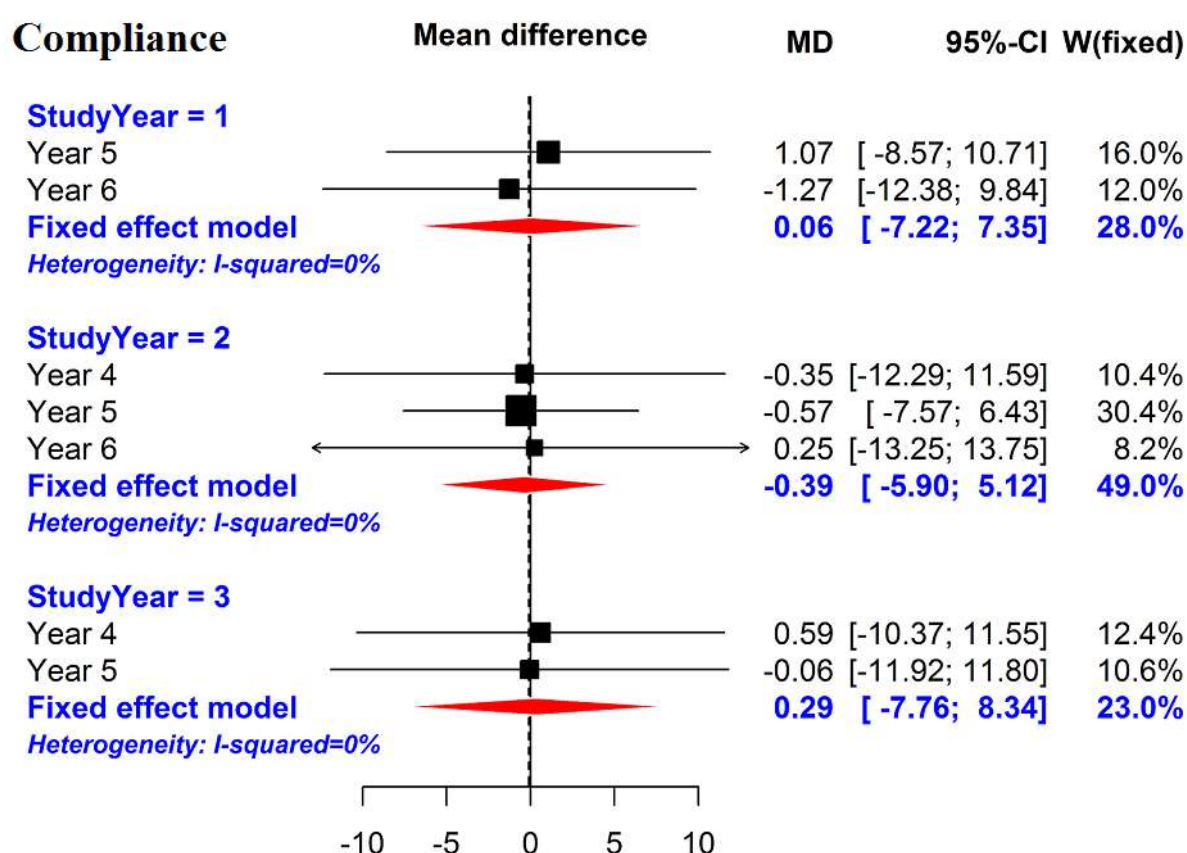


Figure 8. CACE analysis based on asymptotic 95% confidence intervals for pupils with more than 50% compliance

In order to further investigate the implication of compliance in the evaluation of the intervention, we performed CACE analyses for trial and year group with positive intervention effect (shown in Figure 8) using non-parametric bootstrap confidence intervals instead of asymptotic variance approximation. Figures 9 – 11 show the CACE results for Year 5 from the pilot trial, Year 6 from the second year trial and Year 4 from the third year trial. The bootstrap confidence intervals are narrower than in the Schochet and Chiang (2011) approach. A possible explanation is that bootstrapping was done at school level since school is the unit of randomisation; this means the compliance status within school is considered constant across the bootstrap runs. Consequently, it would have dispersed less than a joint bootstrapping done at pupil and school levels.

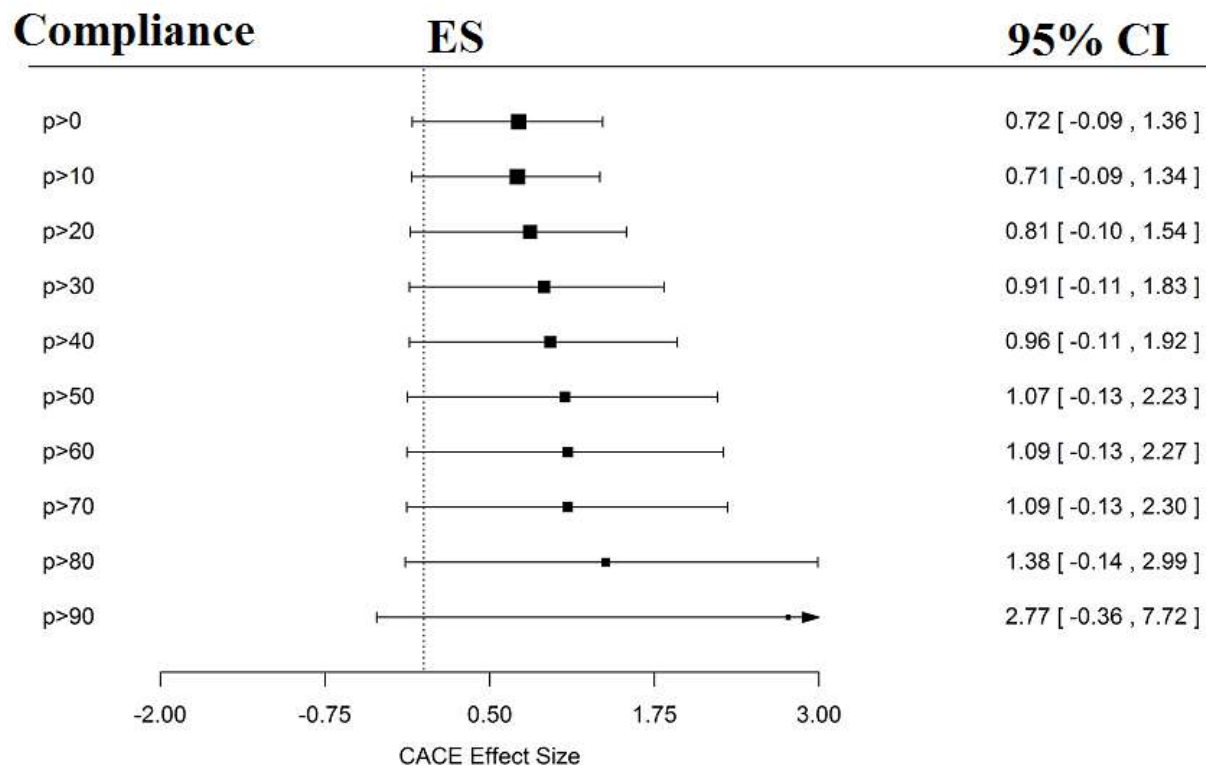


Figure 9. CACE analysis for Year 5 from the pilot trial based on non-parametric bootstrap confidence intervals and a grid of compliance thresholds

Figure 9 further explored the relationship between effect size and adherence for Year 5 from the pilot trial with a positive effect in Figure 8. However, no further exploration was performed for Year 6, which had an overall negative effect because the magnitude of an effect is less important for an ineffective intervention. This plot shows a strong relationship between adherence and effectiveness which reflects the importance of optimum implementation of an intervention.

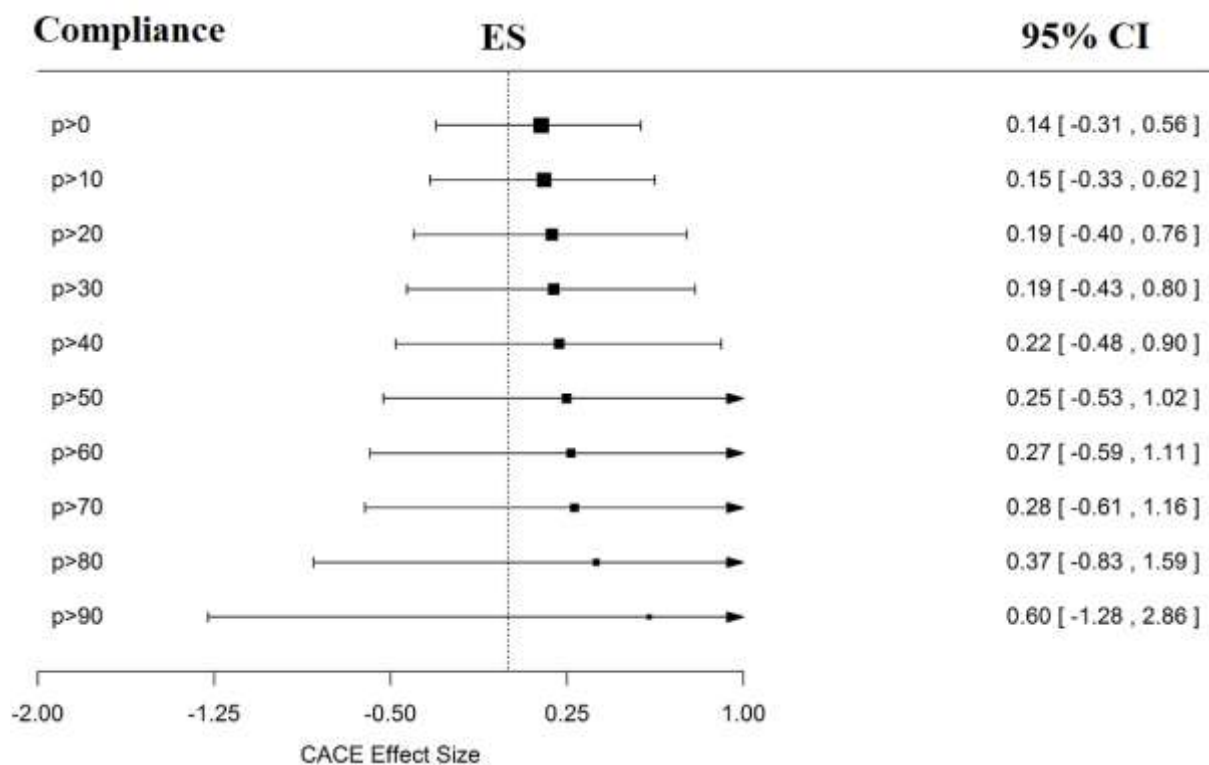


Figure 10. CACE analysis for Year 6 from the second year trial based on non-parametric bootstrap confidence intervals and a grid of compliance thresholds

Similar to Figure 9, Figure 10 further explored the relationship between effect size and adherence for the Year 6 from the second year trial with positive effect size value in Figure 8. This plot also shows a strong relationship between adherence and effectiveness, reiterating the importance of proper implementation.

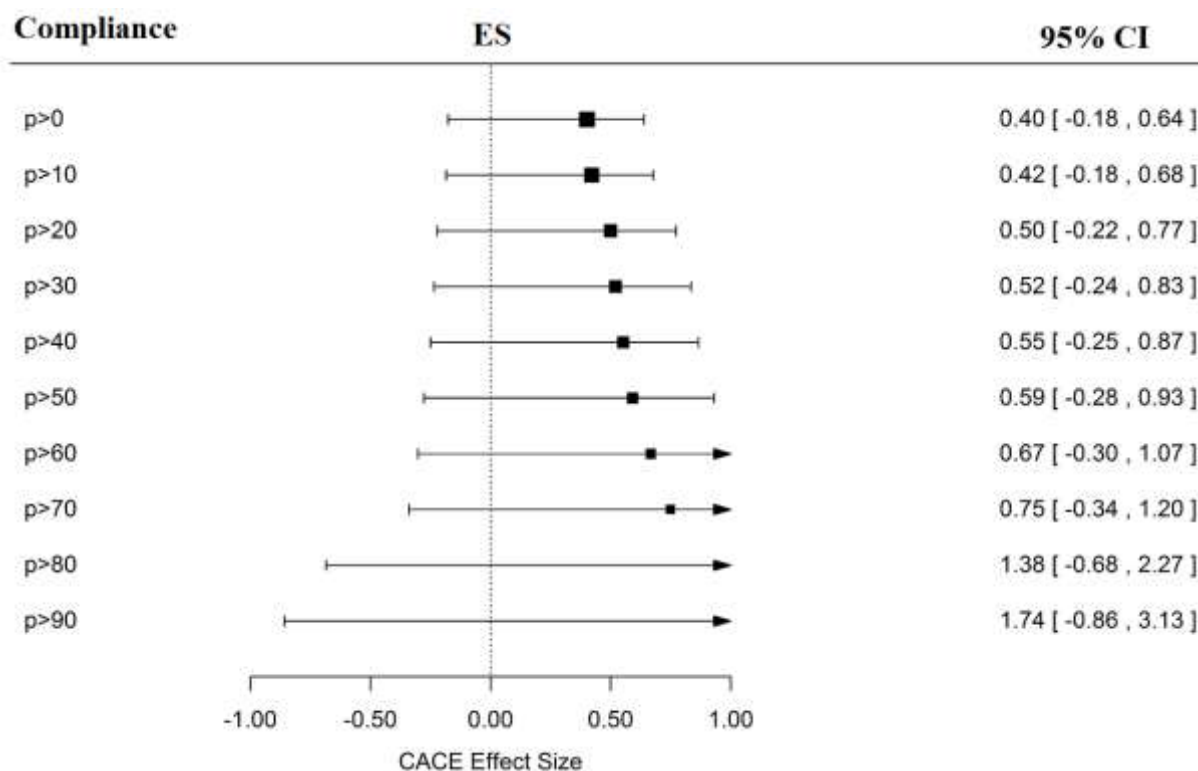


Figure 11. CACE analysis for Year 4 from the third year trial based on non-parametric bootstrap confidence intervals and a grid of compliance thresholds

For the third year trial as well, the relationship between effect size and adherence is further explored for the Year 4 which showed a positive effect in Figure 8, but not for those in Year 5 with a negative effect. Consistent with results in Figure 9 and Figure10, a strong relationship between adherence and effectiveness was demonstrated.

Cost

The annual budget granted by *SHINE* to host schools to run the intervention was £43,600 per year. The grant covered the employment of staff running the programme, start-up resource costs, curriculum resources, extra-curricular visits, end of term celebrations, lunches and a contingency budget. How the grant was to be spent by Intervention Projects was set out in detail by *SHINE* and schools had to report to *SHINE* under each category – actual spends were expected to come in within 10% of the budget. This is reproduced from the Hallé *SHINE* on Manchester: Making it Work handbook (*SHINE*, 2012, p30) in figure 12 below. *SHINE* reported that schools did not deviate much from this budget.

SHINE on Saturday Budget Template

Description	£ week	£ year
Staffing		
Project Management	500	12,500
Teacher x 3	600	15,000
Classroom Assistants x 3	270	6,750
Peer Mentors x 3	30	750
Start Up		
Recruitment and promotion materials	n/a	500
Start up resources	n/a	500
Curriculum		
Resources	30	750
Visiting professionals, transport and visits	48	1,200
Celebration of achievement	n/a	500
Lunches	150	3,750
Maintenance	n/a	school
Contingency	n/a	1,400
Total	1,628	43,600
		£ year budget
Income		
SHINE grant		43,600
School contribution if applicable		
Student fees if applicable		
Other		
Total		43,600

Figure 12. Budget Template for Intervention taken from Hallé SHINE on Manchester: Making it work (SHINE, 2012)

The costs per pupil were calculated in two ways: (1) cost per pupil based on the average number of pupils that attended a project over the third year of the project (35.5). and (2) the costs per pupil as expected by the intervention assuming 60 pupils attend each week.

Table 23. Costs per pupil for the Hallé SHINE on Manchester intervention

Cost based on:	(1) All nominated intervention group pupils (ITT)	(2) Pupils that attended intervention (CACE)	(3)
Cost per pupil	£464	£1228	£727

Process evaluation

Implementation

Piloting issues revealed during first pilot year informing trial

Evaluation procedures

Most stakeholders involved in the project felt that the timeline at the start of the project was too tight. Schools and *SHINE* felt that the timeframe for recruiting staff to the projects as well as nominating eligible pupils and getting pupils through the computerized assessments was challenging. At a discussion midway through the project between the evaluators and *SHINE* it was agreed to give schools longer to do all aspects of the set-up for the main trial years. It was also decided that, based on the uptake rate of nominated children to the intervention, that schools should nominate more pupils for the main trial.

Intervention attendance

Low attendance from the eligible pupils who were allocated to the intervention group was an issue reported by *SHINE* and by the project managers. Projects were generally running at around 35-50 children attending each week compared to expected attendance of 60 children. Attendance has been a key challenge for *SHINE* throughout the evaluation and a number of reasons were reported as barriers to attending *SHINE* for some children, such as other Saturday commitments and lack of some parents' or pupils' engagement with the programme. Various strategies to encourage attendance were used by the project schools (see below in *SHINE* staff feedback section). All children who were nominated in the allocated year group were invited to attend the programme.

Additional pupils from the allocated year group who were not nominated as eligible for the programme (and therefore not included in the evaluation) were also invited to attend to fill up spaces. *SHINE* staff were very keen that schools were making full use of the funding to support as many children as possible. Schools felt that the short time frame at the beginning of the year had meant that they had not had long enough to engage with children and parents to get them to 'buy into' the programme. They also felt that the restrictions of the evaluation (i.e., only one year group was permitted to attend) meant that they were unable to invite pupils who would have otherwise attended, and therefore the evaluation conditions were not fully reflective of normal programme conditions. The *HSOM* project manager said that schools felt 'stuck between a rock and a hard place' with the evaluation restrictions and the demands from *SHINE* that they improve their attendance figures. They did however feel that a longer lead in time for the second year would improve attendance.

Attendance was reported as best from pupils at the host schools. Schools reported that domestic circumstances (e.g., split parent families) and lack of pupil and parental engagement and motivation in the project were reasons for the poor attendance at the projects. It was also reported that other Saturday activities tended to take precedence over attending the *HSOM* programme. The fact that there was not a waiting list and the pressure to maximise numbers meant that it was difficult to enforce penalties for poor attendance (e.g., missing three weeks led to loss of the place). One school reported that girls were less likely to drop out than boys and this self-perpetuated as there was a two thirds bias of girls to boys which possibly made it less appealing for boys to attend. A related issue that was raised was that if partner school staff were not involved in running the project they did not

tend to promote the project to their pupils and some were not supportive in following up poor attendance in those schools.

Intervention reception

Feedback from staff at schools visited and from conversations with the *HSoM* team indicated that the intervention had been received positively by parents, children and by the staff running the programme. Children who attended reported enjoying the programme. Staff and parents reported that children had improved in their confidence during their time attending, particularly in performing in front of others. Parent feedback after the performances at the end of each term suggested that parents felt the programme was worthwhile and that the children were benefitting from it and enjoying it. Schools reported having pupils in other year groups enquiring about the programme and whether they would have the opportunity to attend.

Observations

Pupils seemed to have built up friendly relationships with other children involved and with the teaching staff. The pupils were particularly engaged in the music lessons.

Formal Process Evaluation – second and third years

This section of the report refers to the formal process evaluation phase that took place during the second and third years of the evaluation. It discusses positive features of the delivery of the intervention and some barriers to its successful implementation. As mentioned in the methods section, the analysis is based on data collected through interviews with key stakeholders (i.e., teachers, project manager, Hallé musicians and education director, *SHINE* staff), lesson observations and focus group interview with pupils. This is followed by a summary of key findings regarding the positive features of the delivery and some barriers to the successful implementation of the intervention.

Positive features of the delivery

Pupils' attendance in some schools was good on the day of the visit (for example, 93% of the pupils expected to attend were present on the day of the visit at School 5 and there were 82% of pupils present in School 4).

All schools visited adopted a number of strategies to encourage initial uptake and continued attendance. School 5 introduced a weekly attendance raffle and a competition prize as a way of encouraging pupils to attend. Good communication among the different schools allowed classroom teachers to follow up any pupils with low attendance and also share *SHINE* activities and events with the pupils during the week (Schools 5 and 6). The head teachers' support was perceived as being a key factor in supporting the pupils to attend:

'It is obvious when headteachers are more or less involved in the project or support it more and how it is followed up in schools. Their involvement is really important. One headteacher was fantastic and he always spoke to them about SHINE on Monday and they all attended. It's important that even just once a week it is mentioned in assemblies'. (Project manager, School 6)

Furthermore, the project manager in School 5 volunteered to collect some children from home and take them back at the end of the day to facilitate transport to the *SHINE* school for the parents. A similar initiative took place at School 8 where the project manager provided a minibus service to collect a number of pupils from their primary school on Saturday mornings.

In all the visits, pupils' behaviour was very good, there were very good working relationships among the pupils and with the teachers and there was always a positive and purposeful learning environment in all lessons and workshops. Thematic links were also usefully made in all lessons (for example, the World War 1 theme was explored in the music workshops and the literacy and numeracy lessons in School 5).

Teacher perceptions

All members of staff (teachers, teaching assistants, project managers) interviewed mentioned that the quality of the teaching staff was a key factor for the successful delivery of the intervention. In some cases, teachers and teaching assistants were appointed from the primary school clusters and this was useful due to the teachers' familiarity with the pupils' abilities and needs. Peer mentors were also frequently praised in the interviews because of their commitment to the project and their willingness to support and offer help throughout the project. Last but not least, working with the same musicians throughout the year was identified as an important factor in maintaining pupil engagement through strong working relationships throughout the year. As the following quote illustrates, for instance, the two musicians related particularly well to the children in School 4. Having the same musicians

throughout the year was perceived to be valuable because the relationship had already been established. As the project manager in the school mentioned,

'These two musicians are absolutely spot on. When it went so well last term, I requested that they stay with us cause I love the way they engage the children. Hopefully it will be the same for next year as well. That boy fell in love with some of the music from last term – (the musician) brought a CD and gave it to him'. (Project manager, School 4)

The project managers and teachers in all schools perceived the major outcomes of the intervention for pupils to be an improvement in behaviour and confidence, social skill development and the cultivation of pupils' artistic talents. They felt that all pupils had benefitted from *SHINE* but some pupils had shown more difference than others. Some of the initial behavioural issues were not present or they were very rare. In other words, there was a noticeable improvement in pupils' attitude from the perspective of the teachers and project managers. In School 4, for example, the project manager talked about this attitudinal difference that he had observed in pupils:

'...attitudinally I can see a difference all the time. The headteachers gave me feedback about that which is encouraging and I can tell the teachers here. Some of them that work with the children on a regular basis talk about the massive changes they have seen in school in terms of confidence and about how the children are relating to each other. There is one child whose attitude towards how he relates with others and plays in the playground is just not recognizable, it is such a big improvement for him, and that will set him up before he moves on to high school, and also in terms of being in the classroom and participating and taking risks...'

It was also felt that pupils experience a lot of excitement in lessons because of the more interesting and varied ways in which the material is presented which has encouraged a more open and considerate attitude towards others in the classroom. Members of staff explained that the difference between *SHINE* and children's everyday school experience is that children in *SHINE* are not restricted by being continuously assessed. As the project manager at School 4 mentioned, the children at *SHINE* have the opportunity to practise skills they have been learning in the classroom, draw links with other curriculum areas and see the purpose behind their learning without having the constant pressure of assessment:

'Schools have too much of an emphasis on assessment and in an inner city school where a lot of these children are second language learners... they need to have experiences and intensive teaching before the assessment phase instead of assessment, assessment, assessment. What they get in school on a normal day is like a meal of fish and chips and what we are providing is all the salt and vinegar to make it interesting, exciting and they want more of it. And so we are providing all those creative opportunities to practise the skills they are having in schools... They might be taught the skills of how to do a recount in the classroom. Here they get more of an opportunity to practise it applying it to a particular story they are learning, seeing the links with other artistic areas and then to have an audience for it. Everything they do here has a purpose, it's for the culmination and therefore for an audience.' (Project manager, School 4)

*The children were also given regular opportunities for enrichment through a variety of visits to museums and other educational centres which gave children the chance to learn, experiment and express their viewpoints. The teachers appreciated working in the project as a team where they had a common goal and the opportunity to experiment with more cross-curricular links, have more freedom in planning and work in a fun and relaxing environment. Working with children from many different cultures and linguistic backgrounds added an extra sense of excitement to the teachers' work in *SHINE*.*

'It's really strange but you get here on a Saturday morning and you really enjoy it – it's a reinforcement of what teaching and working with the children really is about! Whereas at school you can worry about their levels, about the planning, that file, that observation, so it's all very focused, you're always pushing yourself, whereas here it's working with the kids.' (Teacher, School 8)

The Hallé musicians

The Hallé musicians interviewed in Schools 7 and 8 talked about the project's success particularly regarding pupils' engagement with their work, their levels of concentration and behaviour, which had noticeably improved since the beginning of the project:

'We do notice quite a bit of improvement from the beginning of the year. I think they are generally more involved and their behaviour does seem better. ... (a particular girl) was impossible at the beginning... she is brilliant now. We have noticed quite a big difference with most pupils that have been coming all year. They seem able to concentrate a bit better and can focus more quickly, now they know what they have to do and can get into it a bit quicker without having to be told'. (Hallé musician, School 7).

Improved behaviour and social skills were perceived as the major outcomes of the intervention from the musicians' perspective. As far as music skills were concerned, singing was perceived as an area where children showed improved attitude and engagement. In particular, it was observed that some children who may not have been comfortable enough to participate by contributing to the group discussion, were happy to take part in a singing activity (musician, School 7). One of the musicians in School 8 made an interesting comment about children being encouraged to develop 'peripheral awareness' as they make music with others in a group as they watch, listen and apply rhythmic awareness 'with someone else doing the same thing. ...that's I think where the strength lies in terms of rhythmic work'. The culmination events were also perceived to be a successful part of the project as the children worked hard towards getting ready to perform in front of an audience putting extra effort and concentration on this task.

Benefits for the children were also mentioned in terms of developing rhythmic skills even though concerns were also expressed about the difficulty in teaching rhythm to each individual child within the limited time available. The musicians in both schools were happy with the support offered by the teachers in the school. The musician interviewed in School 8, in particular, commented positively on the music support offered by two teachers who happened to have musical expertise.

'The fact that they are music support is brilliant cause often, if you get support, it's not music based at all, it's straightforward support and they do what they can, but the progress is limited as a result of that. So having those two teachers specifically has made a big difference on this project I think. ...this is another bonus for us'. (Hallé musician, School 8)

Problems with some pupils' behaviour were perceived as being an occasional barrier to the smooth progression of the musical work that was taking place.

Pupil perceptions

All pupils interviewed reported that they enjoyed going to the Saturday school. They all seemed keen to attend and mentioned that they did their best to attend every Saturday. They were enthusiastic about all the activities that formed part of their day and expressed appreciation for their very good teachers. As one pupil from School 8 mentioned,

'I think we are the luckiest children that we have these teachers in SHINE'.

They described how there were different themes that they were exploring and they enjoyed doing the activities that were relevant to each theme. Pupils in School 5, for instance, mentioned that their theme at the time of the interview was the World War I and the previous theme was about the Roman Carnival.

They particularly enjoyed the opportunities they had to make new friends. Some pupils mentioned that they already had a few friends when they started but, because of carefully organised activities during which they were encouraged to talk to other children they did not know, they quickly managed to make new friends.

Pupils were also keen to talk about the work they did in music, literacy and numeracy and talked with enthusiasm about what they learnt.

In music, pupils in School 5 proudly talked about some musical terms they had learnt:

'You learn about the tempo and the speed of the music, the volume, the pitch...'

'crescendo: getting louder and diminuendo: getting quieter'.

They also mentioned some of the creative work they do in music and appreciated how the musicians give them the opportunity to contribute their own musical ideas to the overall result. They enjoyed how the three different groups come together and perform a whole piece of music which their parents and other people from their schools can attend during the culmination performing events. Pupils' enthusiasm about their musical involvement is illustrated in their comments below:

'It makes you feel like you are actually a part of it, you feel like you contribute what you know to everybody'. (Pupil, School 5)

'In music you're learning terms that you didn't know before and you're learning how music is played and then you're seeing how it feels like to actually be a musician and then you're seeing what you need to do and what you need to know'. (Pupils, School 5)

Last but not least, their attendance in musical events was valued as all pupils were looking forward to the upcoming Hallé *SHINE* concert. Some pupils mentioned their enjoyment in watching the Hallé orchestra perform in the concert hall.

'We watched the orchestra playing about the firebird in our last project. We watched the performance, we had scratch and smell cards, they smelled lovely'. (Pupil, School 8)

The pupils interviewed also talked about their **literacy** work. They mentioned how they were encouraged to discuss with others in groups and share their knowledge, thoughts and feelings. They were also comfortable talking about the previous theme and some of the activities that formed part of that theme. As part of their literacy work, they enjoyed how they were encouraged to use scrapbooks where they could summarise what they had previously done and keep as a record of their learning.

'We have done story writing – we could write any story using our imagination. In the last project, we made a big firebird, we wrote poems on a card, we made cards with a sunflower and seeds in the middle and wrote poems at the back'. (Pupil, School 8)

When they talked about their **numeracy** work, they explained what they had learnt often making links with the overarching theme:

'We did rationing – we made potato cakes which is one of the things that they ate then ...at the same time we learnt about why they ate potato cakes'. (Pupil, School 5)

'Last week for maths we had coordinates and we had to guess where (different) battleships were in the coordinates'. (Pupil, School 5).

'We have done all about maths – additions, subtractions, multiplications, fractions'. (Pupil, School 8).

However, it was observed that in School 7 pupils found it hard to make explicit links with numeracy or literacy and they often needed to be prompted by their teacher who was present during the interview. The following quote illustrates the children's difficulty to talk about their numeracy work. A similar uneasiness to talk about literacy work was also observed in this school.

'Have you done anything with maths, numeracy?

Yeah.

What have you done?

(prompted by teacher) Very often it's with () that you do that.

Yeah.

We were learning about the shipwrecks... and the deserted island...

(pause)

(teacher) And you did codes and things...

The codes were fun... I think it was the Roman...

No, no...

It was morse codes and things.

We had to make all of the secret code, what they put down... that was difficult.' (Pupils, School 7)

Overall, all pupils interviewed felt they were learning in an enjoyable, fun and active way. Their learning was often contrasted with their school work which was perceived by some as being more limited. These points are illustrated in the following quotes:

'You're still learning but it makes it easier to learn... because it's a way that everyone likes to do it – we learn and have fun at the same time.' (Pupils, School 5)

'In school we just have to do work, here we get to do drawing and cooking and we go on trips...' (Pupils, School 5)

'Having extra lessons boosts our learning'. (Pupil, School 8)

'We love SHINE! I want to come to this school instead of my normal school. I want to come to this school every day'. (Pupil, School 8)

Some pupils reported becoming more confident as a result of their participation in the Hallé SHINE sessions and some emphasised enhanced feelings of safety.

'Coming here and making new friends gives you the confidence in whatever you do.' (Pupil, School 5)

'I made new friends, I am confident here.' (Pupil, School 7).

Disappointment for the end of the project was also expressed by a number of pupils:

'We are sad because after this project, we are not going to do SHINE any more'. (Pupil, School 8)

Lesson observations

In the literacy lesson in School 5, the teacher made good use of questioning using children's prior knowledge on the theme under investigation. The children worked in small groups to brainstorm ideas and they were then keen to share ideas with the whole group. The lesson was well-structured with clear explanation of terminology and relevant vocabulary. Children were encouraged to participate and express their views with their peers. They were also involved in a purposeful writing activity and were then encouraged to read out examples of their writing. Some children demonstrated good writing skills as evidenced in the quality of work that was produced.

The music workshops observed were a good example of the engaging work that was taking place in the project. There were some excellent features of pupils' musical work. In School 5, the pupils were given the chance to contribute to the musical result as teachers were picking up on their ideas and they were taking these ideas forward. There was effective questioning throughout the lesson. The teachers were very encouraging of children's responses and there were also opportunities for children to have control with some children volunteering to use simple conducting techniques to act as leaders for the group. In addition to the two Hallé musicians, there were also two teaching assistants who were fully involved throughout the session. Musical terminology was used effectively. It was apparent that the children were familiar with musical terms as they had learnt them before and they were using the terms accurately during the session.

Overall, there was a high level of participation as all children were involved in the group activity and seemed engaged. There was good group work as the pupils worked with the musicians and each other. The teachers offered lots of opportunities to the children to contribute their ideas and were responsive and encouraging. These ideas were then used constructively by the musicians and were incorporated into the musical task. The behaviour was excellent and the atmosphere was relaxed, encouraging and fun. There were positive and constructive relationships between musicians and children.

Another music session observed (School 8) was more exploratory as it was the first workshop in a particular project with the theme being introduced to the pupils. This was an active session where pupils were encouraged to think of and share their ideas regarding the theme. They then made decisions using visual imagery about aspects of their story which led to singing and a preliminary composing activity devising a short motif using particular notes.

Another feature of good practice observed was the musicians' high expectations of the pupils' work in School 7. They were expecting more and were reinforcing this through constant encouragement and constructive feedback.

Culmination events

The two culmination performing events demonstrated the effort and hard work pupils, musicians and teachers had put into the development of the project in previous weeks. Practice sessions preceded the final performance and their purpose was to finalise the

musical piece, remind pupils of their role to the piece, help them improve their musical contribution and add the final touches. All children participated and were generally very well focused on task and responded well to the musicians. There was good motivation and engagement with the pupils showing enjoyment of the music activities and performance rehearsal.

School 6

The majority of the musical input and ideas was coming from the musicians. This would be expected as all musical parts (instruments, sounds, timings, dynamics etc.) were jointly agreed in previous sessions and they were written in order on a flipchart which was then used as a reminder during these practice sessions. During the day, teachers, teaching assistants and peer mentors were busy decorating the school hall with the work that the pupils had completed as part of the project. Pupils' work was artistic, imaginative and exemplified aspects of their literacy and numeracy work.

The children seemed very positive and happy to take part in the performance. The musical result was very good. Overall, appropriate sounds were used which represented each thematic idea well, musical elements (such as dynamics and tempo) and musical devices (such as repetition and variation, development of ideas, use of instruments and sound effects) were used appropriately showing imagination, originality and variety in the final musical composition. The audience were engaged and parents seemed enthusiastic with the musical result and their children's role in it.

After the musical performance, some pupils gave a short speech following a written script that they had prepared. There were also video clips of pupils who talked about their dreams and aspirations for the future. They talked about some possible steps they were going to take to achieve their dreams. In the meantime, there was variety of dance performances, such as line dancing among others.

At the end of the whole performance, certificates were given out to all pupils. Those pupils that had exceptional attendance were particularly rewarded. Pupils seemed to have a real sense of achievement at the end of the day. They were happy to show their certificates and share their achievements with their parents.

School 4

The music performance was from memory and the focus was on rhythm which was often quite challenging musically. The activities promoted learning in music appreciation and production and pupils made progress during the rehearsal. There were very good levels of pupil engagement and opportunities for the pupils to work together, with many practical elements to the learning. The children worked well together to achieve the created musical interludes. The musicians were enthusiastic about the music and about the pupils' performances of it. The musicians had a very nice manner with the pupils and established and maintained a good working relationship between themselves and the pupils.

The teachers' involvement in the practice sessions was varied. When the teacher played an active role in the performance rehearsal, this led to greater focus from the children. Teaching assistants helped alongside the pupils, organising but also taking an active role, helping and guiding, and responding well to individual needs.

All children seemed to enjoy the culmination performance to parents and guests.

Barriers to successful implementation - Suggestions for future development

Pupil attendance was often low and this was considered to be a barrier to the successful implementation of the intervention. For example, low attendance was observed in Schools 6 and 8 on the day of the visit. Low pupil attendance was often mentioned by teachers as a barrier that they were trying to overcome:

'I think for us that's the big difficulty, having parents who are not engaged – those kinds of cultural barriers are quite difficult for us. ...The kids really enjoy it, but they need their parents to bring them here.' (Teacher, School 8)

In lessons, even though the level of pupil engagement and participation was generally very good (as mentioned above in the *lesson observations* section), a more critical dimension could have been added in some of the activities so that children would have had more opportunities to reflect on, correct or improve their work. Further challenge and more constructive feedback could help pupils make more progress with their work. These observations were made in a literacy lesson (School 5), in a music workshop (School 8) and in a practice session before a culmination performing event (School 4).

A particular barrier regarding recruitment was mentioned by the project manager in School 7. Problems with recruitment seemed to arise because of pupils' lack of availability to attend *SHINE* on a Saturday or because of a possible lack of communication with schools and parents about what *SHINE* would offer. As the project manager mentioned,

'People had different ideas about what we were offering... there has been an imbalance between the commitment from the schools. ...Schools need to be more informed, you do really have to step up and really commit to that. Maybe schools didn't quite realise what was going on.' (Project manager, School 7).

Overall, pupils' experiences of *SHINE* were very positive and this was obvious throughout the five formal visits as part of the process evaluation. Pupils seemed to be experiencing a sense of belonging, having a good time and were kept busy throughout the day. They also showed good behaviour and satisfactory levels of participation. They had the opportunity to experience good music teaching and be creative through their involvement in often elaborate pieces of artwork in preparation for the culmination performances. They gained a sense of confidence in performing in front of an audience which resulted in an obvious sense of achievement and personal satisfaction at the end of each performance. They were also encouraged to develop an interesting view towards 'core subjects' such as literacy, numeracy and science as these were taught in a thematic way following the theme that was being explored. This led to an understanding of relevance and added a sense of purpose to many of the activities in which the pupils were involved. However, on the basis of the limited evidence collected regarding the exact nature of the teaching of these 'core subjects', the literacy, numeracy and science lessons observed did not seem to engage the learners sufficiently cognitively and critically. A further level of challenge could push learners to think harder and be cognitively engaged at a deeper level.

Hallé education director and Hallé programme manager

An interview with the Hallé education director was conducted during the visit in School 4 and a further longer interview was conducted separately with the Hallé education director and the Hallé programme manager at the Hallé offices. Interviews explored their thoughts on implementation, what went well, relationships and future directions

The Hallé education director praised the work of the *HSoM* manager whose personality and good working relationship with both the musicians and the teachers helped inspire more confidence to the musicians and helped schools improve their practice by a hands-on and

consultative approach. Furthermore, it was mentioned that teachers' initial reticence and nervousness around the musicians gradually subsided as better working relationships were formed through improved communication and familiarity. This was facilitated by trying to keep the same musicians in each project. This improved partnership working between the teaching staff and the musicians was considered to be a key factor in the successful implementation of the project:

'That has also been helped by keeping the musicians in each cluster the same which wasn't the plan at the beginning because what we were trying to do was have a different instrument each time. We have now kept the musicians the same in each cluster and what we found is that they got to know people and that's been beneficial for all'. (Hallé education director)

Additionally, the music work that took place in the schools as well as children's concert attendance allowed children to develop greater musical awareness and appreciation. Through children's practical involvement in music making, some children with musical potential were also identified and offered further musical training by the local music hub.

The Hallé identified a number of benefits of the project from their perspective. The longitudinal aspect of the project gave the players the opportunity to understand the children's perspective and wider issues. This helped the musicians to develop their practice and skills as well as building relationships with the children over the course of the project. It also provided positive role models for the children, and gave them the opportunity to attend events that they may not have experienced otherwise – with pupils attending a concert at the Bridgewater Hall and seeing the orchestra performing, which in some cases was the children's first visit to Manchester.

The relationship between the Hallé and *SHINE* is perceived to have worked very well during this project and that they are now also collaborating outside of this.

Scheduling the musicians was reported as one of the challenges for the Hallé as Saturday is concert day and musicians were not always available. Recruiting pupils to the intervention and encouraging continued pupil attendance was another challenge. Good communication between staff in a project and between the project and the partner schools was seen as an enabling condition to the intervention working well and schools being able to recruit enough pupils. The Hallé also perceived that the limitations of the evaluation made it difficult for projects to recruit pupils. The nomination process and assessments were thought to have held up recruitment time in the project and they had had reports from staff that schools and some parents were dissatisfied with the requirement for pupils to be excluded where they formed the control group.

Hallé players' feedback

During the course of the evaluation as part of the observation visits a number of informal conversations took place between the musician supporting the project and the evaluation team. In addition more formal (i.e., semi structured) telephone interviews were carried out with two of the musicians. These interviews were arranged by the Hallé project manager with the contact details of the musicians being passed to one of the research team. The notes below are based on both the more formal telephone interviews and informal conversations. However, given both the size and nature of the sample (i.e., a convenience sample) they should be viewed as more illustrative of the project, rather than a robust and representative reflection of the musicians' views.

Throughout the project period we were impressed by both the enthusiasm and expertise, of the musicians taking part in terms of teaching and working with the pupils. There was a

general response that they saw this as valuable work, both in terms of the Hallé mission as well as more generally for helping the children. As was pointed out on several occasions the orchestra has a long history of supporting the Manchester wide community, and in particular the more disadvantaged members of that community, and more specifically in terms of music to help provide access to high quality musical experiences.

In all of the instances the musicians had previous experience of teaching and supporting music in the community, for example, through the 'Adopt a Player' scheme. Whilst they pointed out that the pupils starting point in terms of music was generally very low, this for them was not surprising, and in all instances felt that worthwhile progress was made during the 20 or so week period. Indeed as one of the interviewees commented, any progress even if simply being able to follow a rhythm 'has' to be of some benefit. This said, all of the players were able to identify at a least a few pupils who made very good progress and a large majority who they felt were enthusiastic and committed to the music element of the project.

Generally the musicians felt supported by the teachers and very much enjoyed working with them. Besides this they felt that within reasonable expectations for this kind of work the sessions were well organised and controlled, and any problems or issues effectively dealt with. They said there were not many drop-outs, but also that having new pupils joining part way through was not always easy and could create difficulties. Generally, they felt the behaviour of the pupils was reasonable, and was certainly not for the vast majority a barrier to their progress.

From discussions on the broad aims of the project as well as more specifically the theory of change, it was felt that the core elements had been, or could have been, met. In all cases it was also felt that the project highlighted the need for more and better music teaching in mainstream education, and that they would like to see such projects and initiatives continue in the longer term.

SHINE staff feedback

SHINE perceived the intervention as successful in Manchester and believe that based on their tracking data, the majority of pupils regularly attending *SHINE* are making two or more sublevels of progress during the year which is similar to the London *SHINE* on Saturdays programme. Schools have reported to *SHINE* an impact on pupils' social and emotional skills as well as their behaviour and confidence.

SHINE felt that the programme works for a number of reasons. It gives disadvantaged or struggling pupils a different experience of school than they would otherwise experience. During Saturday sessions they have more opportunity to speak to their teachers about issues they might be experiencing and get more attention. This has enabled them to build really good relationships and opened up conversations for children to talk about where they are struggling. It also gives some children the chance of doing something successfully at school particularly when children are struggling or disengaged with week day school. It has also given pupils and schools the opportunity and access to enrichment activities that they would not normally have accessed including visiting the Hallé. The smaller class sizes were seen as a positive feature of the programme particularly for teaching core skills; however, there is a trade-off between the cost of the intervention and how small the class size could be.

SHINE also reported that the intervention has a big impact on the teachers delivering it. No projects have had any issues with recruiting teaching staff or programme leads despite the fact that for teachers it usually means a six day working week. Feedback to *SHINE* has been that working at *HSoM* on a Saturday allows teachers to be more innovative in terms of lesson delivery and curriculum. The mix of teachers from different partner schools has been a

positive feature and has allowed better relationships to form between those teachers and between the schools.

They report that schools have committed fully to the programme and all schools involved were intending to apply to *SHINE* for further funding to continue offering the programme. Some schools were also putting up some of their own money towards the costs. Local authority music support has been valuable to projects in setting up.

From their position, where the projects have worked well it seems to be where there is a functioning partnership of schools which are pro-active in chasing up things (e.g., attendance), take an active interest in the curriculum and that value the programme during the week days at school. The least successful projects have not had good communication systems in place between the schools involved or the links with week day school. All schools have bought into the programme but the leadership and organisational ability to deliver has varied and lead to variability in delivery of programmes. A strong project manager along with high quality staff members to deliver the sessions were also important for good delivery of the programme. Both of these aspects, in turn, lead to the creation of good experiences for the pupils who attend the programme. To get the right project manager requires an investment of time for *SHINE* and the school head teacher to identify the right person. Only one school was delayed in recruiting a project manager. In most of the projects the project manager was an experienced teacher with a passion for music but not necessarily an existing staff member within the host school; this was different from the London model where most project managers were senior leadership staff in the host school.

A key challenge for *SHINE* has been attendance at the projects throughout the evaluation. It is important to *SHINE* that the maximum number of children are able to benefit from the grant money. Reasons that nominated children did not attend were reported as: other Saturday commitments e.g., music, sport, religious activities, family arrangements; or a lack of engagement with the programme either from parents or from pupils themselves particularly if it is seen as selecting those that are struggling. The selective nature of the programme has also caused an issue when promoting the project in school as some children who would like to attend were unable to do so. Where attendance was good, strategies used were: visits and promotion of the project across all cluster schools through staff and parent meetings. For example, one project had sent weekly newsletters with a section to return the following week; some projects incentivised attendance with prizes and rewards. These strategies were on top of regular communication strategies with parents e.g., texting parents on Thursday/Friday before sessions, following up missing pupils through phone calls. Attendance was generally higher for pupils in the host schools than for other schools.

The evaluation restrictions of only including those pupils who had been nominated by teachers in advance of the project had been more restrictive than *SHINE* initially believed it would be. It had been difficult for teachers to predict which pupils would attend the intervention at nomination. Not being able to speak to parents in advance of the launch of the project had been perceived by schools as another barrier to recruiting pupils to the intervention. The additional evaluation requirements of nomination and assessment at the beginning of the year added to logistic challenges within schools that distracted from the initial project set-up.

Summary of positive features of the delivery

- Pupils' behaviour was very good and positive working relationships at all levels were evident throughout the process evaluation visits.

- The quality of the teaching and support staff (project manager, teachers, teaching assistants, peer mentors, musicians) was perceived to be a key factor for the successful delivery of the intervention.
- Key stakeholders (project managers, teaching staff and musicians) perceived the major outcomes of the intervention for pupils to be an improvement in behaviour and confidence, enthusiasm and focus in learning, social skill development and the cultivation of pupils' artistic talents, including rhythmic and singing skills.
- All pupils expressed excitement and enthusiasm about different parts of their engagement in the Saturday school, including the chance they had to make new friends, their musical involvement and the new learning opportunities they were offered to make progress in literacy and numeracy in a fun and active learning environment.
- Active pupil participation was evident throughout the lessons and workshops observed. Pupils felt comfortable to express their views and purposefully work with others in groups.
- The culmination performing events provided a good opportunity to showcase the effort, enjoyment and hard work pupils, musicians and teachers had put into the development of each project.
- Teachers' involvement in the music practice sessions was varied. When teachers took on a more active role in the performance rehearsal, children seemed more focused and engaged.
- Teaching assistants helped and guided pupils responding well to individual needs.
- A number of strategies adopted by schools helped encourage initial uptake and continued attendance.

Summary of barriers to successful implementation

- Low pupil attendance was a barrier to the successful implementation of the intervention.
- The evaluation requirements of pupil nomination, recruitment and assessment at the beginning of the year were challenging for the schools and caused difficulties in the initial project set-up.
- The least successful projects have had poor communication systems in place with the schools involved.
- Pupils' poor behaviour occasionally caused difficulties but most of these were overcome with time as pupils made progress and gained in confidence.
- Further challenge and more constructive feedback in 'core subjects' could help pupils make more progress with their work by engaging them at a deeper cognitive level.

Conclusion

This report provides the results from a one-year pilot efficacy and two-year main efficacy trial, with impact and process evaluations. The trial was designed, conducted and reported in a rigorous fashion and, despite some limitations due to attrition and missing data, as such should be used for interpretation and to make conclusions and/or recommendations. The impact evaluation results were consistent across all outcomes and age groups with no statistically significant effect in one direction or another. Therefore we can conclude that for all educational and wider outcomes that we measured *HSoM* is unlikely to be a cost-effective intervention. In contrast, the process evaluation findings revealed a very positive picture of pupils' involvement and engagement in their Saturday school activities. Very good working relationships among the pupils and with the teachers were observed and there was always a positive and purposeful learning environment in all lessons. Key stakeholders felt that pupils were making noticeable gains in behaviour, confidence and the development of their social skills.

Impact evaluation: Strengths

- A robust RCT design with rigorous methods to minimise all potential sources of bias; three years' impact data (one year pilot trial and two years' main trial);
- Robust trial conduct and reporting to minimise all potential sources of bias;
- Reasonable compliance to the intervention, with the mean number of sessions attended between 14 and 15 out of a possible 23 to 25 sessions (around 50%); and
- A range of educationally significant outcomes measured.

Process evaluation: Strengths

- All five projects of the main trial visited as part of the process evaluation;
- Quality assurance of the observations carried out at the beginning of the site visits, leading to enhanced confidence in evaluative judgements made in subsequent visits;
- A range of data collection methods used during the visits ensuring a holistic picture of key stakeholders' perceptions and nature of pupils' work.

Impact evaluation: Limitations

- Consistently *reasonable* uptake on the offer of a place on the intervention and sub-optimal attendance at the intervention through the three years of implementation may have led to the potential for dilution of effect in the outcomes;

- Attitude outcomes showed huge variability between pupils of the same school; such variability may be an indication of lack of reliability in the set of items used to quantify the scores or a scale issue since the attitude outcomes are on a different scale from the other outcomes;
- A reasonable amount of missing data from the InCAS assessments during the second and third years of the trial (much of the missing data due to technological issues with administering the assessment: data missing for this reason is likely to be *at random* across the schools and would affect both intervention and control group pupils similarly);
- During the first and third years of the trial one school (in each year) invited pupils from the control year group to attend the intervention (pupils from the intervention group were also invited to attend); pupils from the control year group who had taken up the intervention were allowed to continue; however, no follow-up of attendance or encouragement of those who hadn't attended was done for control group pupils.

Process evaluation: Limitations

- Cross-sectional design of the process evaluation did not allow an in-depth follow-up of issues raised in each project in a systematic way;
- Pupils were selected to take part in the focus interviews by their teachers, so we cannot be certain that the sample of pupils interviewed reflected the whole range of views and attitudes towards the intervention;
- Not always possible to observe full lessons (literacy, maths or music) during the process evaluation visits; therefore, informed judgements about the extent to which intended learning objectives were met could not be made in all instances.

Interpretation

This study found that there was no impact of being nominated for the *HSoM* programme on attainment in reading or mathematics and no impact on music or attitudinal outcomes. This finding is in line with the existing evidence: meta-analyses of evaluations of existing out-of-school programmes found small or no effects on literacy or maths achievement when high quality robust research designs and methods were used (e.g. Kidron and Lindsay, 2014; Lauer et al., 2006). Kidron and Lindsay *did* find a small effect on academic outcomes when qualified teachers were used (as is the case in this evaluation); however, this study did not confirm this finding in this intervention. Another EEF funded pilot evaluation of a similar *SHINE* programme with pupils in Year 7, *SHINE in Secondaries*, undertaken by this evaluation team, also found only very small insignificant effects of the programme on numeracy and literacy (Menzies et al., 2015). Therefore, on the basis of consistent evidence from four evaluations of this intervention we would not recommend the funding of future research to further evaluate the intervention.

Although there were problems with non-compliance, with only about 50% of pupils getting the intervention as intended, our analysis (CACE) that took this into account did not materially alter the picture of no effect of the intervention. This finding is important and demonstrates the importance of evaluating novel educational interventions using rigorous experimental evaluation designs.

References

- Chamberlain, T., Aston, H., George, N., Sharp, C., and Sanders, D. (2011). *Evaluation of SHINE on Saturdays*. National Foundation for Educational Research.
- Centre for Evaluation and Monitoring. (2016). *CEM.org website - School Assessments: Primary*. Retrieved 9 March 2016, from <http://www.cem.org/primary>
- Cook, T.D. and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings*. Rand McNally, Chicago, Illinois.
- Crawford, S.T. (2011). *Meta-analysis of the impact of after-school programs on students reading and mathematics performance*. Denton, Texas. UNT Digital Library.
- Fashola, O.S. (1998). *Review of extended day and after-school programs and their effectiveness. CRESPAR report no. 24*. Center for Research on the Education of Students Placed at Risk.
- Harlow, K., & Baenen, N. (2001). *The effectiveness of the Wake Summerbridge Summer Enrichment Program: Eye on evaluation: E&R report*. Raleigh, NC: Wake County Public School System, Department of Evaluation and Research. (ERIC Document Reproduction Service No. ED 466 489)
- Haywood, S., Griggs, J., Lloyd, C., Morris, S., Kiss, Z., and Skipp, A. (2015). *Creative Futures: Act, Sing, Play. Evaluation Report and Executive Summary*. Education Endowment Foundation, London: UK.
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32(4), pp. 341-370.
- Kidron, Y., and Lindsay, J. (2014). *The effects of increased learning time on student academic and nonacademic outcomes: Findings from a meta-analytic review* (REL 2014–015). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Kokotsaki, D. (2015). *Improving the primary-secondary transition in music education*. Research project funded by the Nuffield Foundation. Available at: <http://www.nuffieldfoundation.org/sites/default/files/files/Dr%20Dimitra%20Kokotsaki%20-%20Dec%202015.pdf>
- Kokotsaki, D. (2016, in print). Pupils' attitudes to school and music at the start of secondary school. *Educational Studies*.
- Lauer, P., Akiba, M., Wilkerson, S., Aphorp, H., Snow, D. and Martin-Glenn, M. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76(2), pp.275-313.
- MacBeth, J., Kirwin, T., Myers, K., McCall, J, Smith, I., and McKay, E. (2001). *The impact of study support (Research Report RR273)*. Department for Education and Skills.

- Martinez, A., & Cosentino de Cohen, C. (2010). *The national evaluation of NASA's Science, Engineering, Mathematics and Aerospace Academy (SEMAA) program*. Cambridge, MA: Abt Associates.
- Menzies, V., Collyer, C., Kokotsaki, D., Hewitt, C., Younger, K., Wiggins, A. and Torgerson, C. (2015). *SHINE in Secondaries. Evaluation Report and Executive Summary*. Education Endowment Foundation, London: UK.
- Ofsted (c2012-2106). *Ofsted Data Dashboard*. Retrieved 9th March 2016, from <http://dashboard.ofsted.gov.uk/index.php>
- Scott-Little, C., Hamann, M., and Jurs, S. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, pp.387–419.
- Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Open Med*, 4 (1), pp. 60-68.
- Shadish, W., Cook, T. and Campbell, D. (2001). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- SHINE (2012). Hallé SHINE on Manchester: Making it work. Unpublished project handbook.
- Schochet, P. Z., and Chiang, H. S. (2011). Estimation and identification of the complier average causal effect parameter in education RCTs. *Journal of Educational and Behavioural Statistics*, 36 (3), pp. 307-345.
- Standley, J. (2008). Does Music Instruction Help Children Learn to Read? Evidence of a Meta-Analysis. *Update: Applications of Research in Music Education*, 27(1), pp.17-32.
- Torgerson, D. J., and Torgerson, C. (2008). *Designing randomised trials in health, education, and the social sciences: An introduction*. New York: Palgrave Macmillan.
- Torgerson, C. J., Wiggins, A., Torgerson, D. J., Ainsworth, H., Barmby, H., Hewitt, C., ... and Tymms, P. (2011). *Every child counts: The independent evaluation*. (Research Report DFE-RR091a). Department for Education. London: UK.
- Tymms P.B. and Merrell C. (2003). *Screening and interventions for inattentive, hyperactive and impulsive young children*. Paper presented at the Evidence Based Policies and Indicator Systems Conference, London, July.
- Welch, G., Saunders, J., Hobsbaum, A. and Himonides, E. (2012). *Literacy through music. A research evaluation of the New London Orchestra's Literacy Through Music programme*. International Education Research Centre (iMerc). London: UK.

Appendices

Appendix A: Evaluation information sheet for schools (recruitment meeting)

Appendix B: School consent form

Appendix C: Parent information/opt-out letter

Appendix D: Parent consent form (Pupil focus group)

Appendix E: Analysis plan

APPENDIX A: Evaluation information sheet (recruitment meeting)

SHINE on Manchester with the HALLÉ Orchestra

Evaluation information sheet

[Note: this information sheet will be used at the school recruitment meeting, September 10th 2012]

Hallé SHINE on Manchester

Evaluation information sheet

The Education Endowment Foundation has asked researchers at Durham University to evaluate Hallé SHINE on Manchester, to find out how well SHINE works in improving outcomes for the children who take part. The researchers at Durham University will not change how SHINE is delivered in Manchester. They have designed a study that will measure improvements in literacy, numeracy and attitudes for the children who take part and will compare these with the same outcomes for children who have not taken part. All recruited schools will take part in SHINE, some with Year 5 children and some with Year 6 children.

2012-13: Pilot Trial

By September 2012 three hub and at least 9 partner schools will be recruited to Hallé SHINE on Manchester and to the evaluation. The hub schools will receive the award of a grant to deliver SHINE to children from the hub school and their partner schools. The schools taking part in the first year of the evaluation will be part of the 'pilot trial'.

The process for recruitment to SHINE/evaluation for all schools will involve signing a memorandum of understanding with the evaluation team to:

- agree to identify up to 20 children in both Year 5 and Year 6 using a consistent approach;
- agree to random allocation of ONE of these year groups to SHINE (and the other group will be a control group);
- agree to baseline testing using computer-based assessment provided by the evaluators with all Year 5 and all Year 6 children; and
- agree to outcome testing using computer-based assessment provided by the evaluators with all Year 5 and all Year 6 children.

All recruited schools will be able to offer Hallé SHINE on Manchester to *up to 20* children in Year 5 **or** *up to 20* children in Year 6.

At the September information/recruitment meeting the evaluation design and the study procedures will be explained in detail, and there will be an opportunity to ask questions. Following this meeting each school will identify up to 20 children in Year 5 and up to 20 children in Year 6 and will give this information to the Durham researchers.

Durham University will provide schools with the necessary software, information and support to administer the InCAS assessment. InCAS is a computer-based assessment provided by the Centre for Evaluation and Monitoring (CEM). It includes modules on numeracy, literacy, developed ability and attitudes. Schools will complete the assessment with all children in Year 5 and Year 6 and return the data to CEM. Schools can then download the assessment results for their own use.

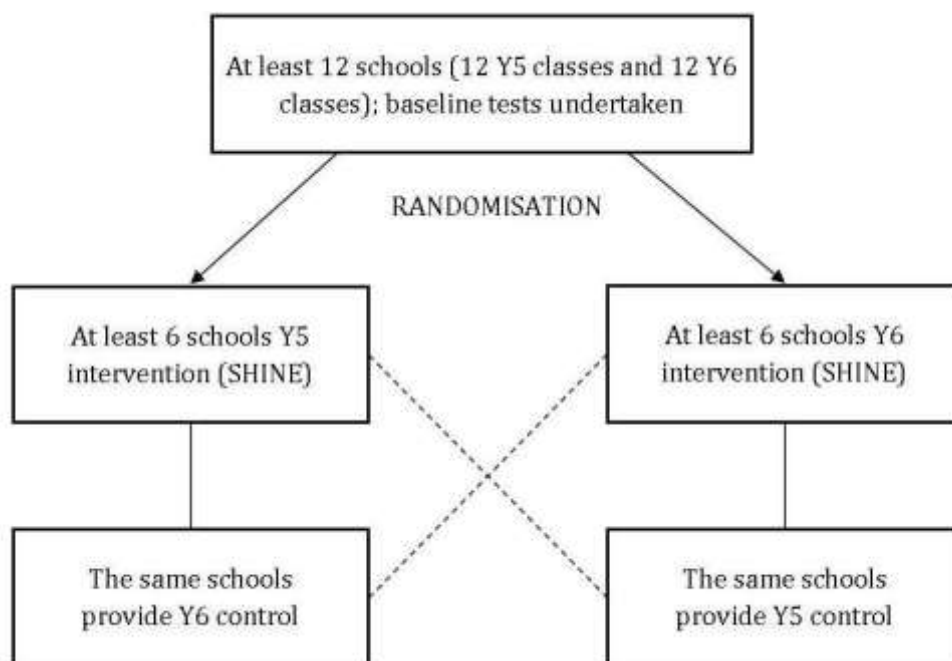
Schools will then be *randomly allocated* (like in a lottery) half to offering SHINE to Year 5 children and half to offering SHINE to Year 6 children. The Durham researchers will inform SHINE of the allocation and SHINE will inform each school which year group will be offered the intervention. Only intervention children and parents will be informed (not control children). Consent will be by opt-in for the intervention children and opt-out (for data use) for the control children. All pupil data will be analysed anonymously and will be kept completely confidential.

Oct. 2012 - June 2013 – Hallé SHINE on Manchester will go ahead with no interference from the Durham researchers, although they will come and observe some of the activities.

In July 2013 outcome testing (InCAS assessment) will be undertaken by schools with support from Durham University as before, with all Year 5 and Year 6 children.

Design

To learn about how well SHINE works we will compare children in both year groups. We will compare the average outcomes of Year 5 children who have received the intervention with the average outcomes of those in the Year 5 control classes. We will also compare the outcomes for Year 6 intervention children with those in the Year 6 control classes.



Year 5 schools in the pilot trial will continue with the intervention and control conditions in Year 6 in 2013-14 to provide one-year follow-up data.

Pilot trial schools will then run the intervention for Year 4 or Year 5 children in 2013-14 using the same design as above. The data from this year will be added to the data from 2012-13 in a meta-analysis, where the data from each year are combined to give more reliable results.

2013-14: Main Trial

In September 2013 at least 20 further schools will join the trial: 5 hub schools and at least 15 partner schools. These schools will take part in the assessment, randomisation and intervention in just the same way as the pilot trial (which will mean that ultimately, we can look at all the data together for greater certainty about the outcome).

Schools in the pilot trial will end their participation in the evaluation after 2013-14; schools in the main trial will continue for a second year in the same way as the pilot trial schools, ending their participation in the evaluation after 2014-15.

Contact Details

Schools are welcome to contact the Durham University team for more information about the evaluation and assessment:

Professor Carole Torgerson, Durham University, Principal Investigator, Independent SHINE evaluation

Contact: carole.torgerson@durham.ac.uk; School of Education, Durham University, Leazes Road, Durham, DH1 1TA. Tel.: 0191 334 8382

Victoria Menzies, Trial Coordinator

Education Endowment Foundation

Contact: **victoria.menzies@cem.dur.ac.uk**, CEM, Mountjoy Research Centre, Durham University, Stockton Road, Durham, DH1 3UZ. Tel.: 0191 334 4177

Kirsty Younger, Trial Administrator

Contact: **kirsty.younger@cem.dur.ac.uk**, CEM, as above. Tel.: 0191 334 4176

APPENDIX B. School Consent Form

Independent Evaluation of SHINE on Manchester Saturday Schools

☐ I confirm that I have read and understood the information sheet for the above evaluation and have had the opportunity to ask questions;

☐ I understand that all children's results will be kept confidential and that no material which could identify individual children or the school will be used in any reports of this evaluation;

☐ I agree to identify up to 20 children in both Year 5 and Year 6 to be eligible to be offered the SHINE intervention;

☐ I agree to random allocation of ONE of these year groups to be offered the SHINE intervention (and the other group to be a control group);

☐ I agree to baseline and outcome testing using computer based assessment provided by the evaluators with ALL Year 5 and all Year 6 children. Opt-out for individual children's data to be used in the evaluation will be offered to all parents.

☐ I consent to the school taking part in the above study.

Name of headteacher

School

Tel no

Email address

Signature of headteacher.....

Date.....

Thank you for agreeing to take part in this research. Please return this consent form at the information meeting or afterwards by post to:

Kirsty Younger, Trial Administrator

CEM, Rowan House, Mountjoy Research Centre, Durham University, Stockton Road,
Durham DH1 3UZ

Appendix C. Parent information/opt-out letter



School of Education
Durham University
Leazes Road
Durham
DH1 1TA

19th January 2015

Dear Parent or Guardian,

RE: Investigating the impact of Hallé SHINE on Manchester on educational attainment

I am writing to let you know about a study which is taking place in your child's school this year (2014/2015). The Education Endowment Foundation (EEF) is funding the study to find out how much the Hallé SHINE on Manchester Saturday school programme helps children improve their English and maths skills. The study is being carried out by Durham University.

The head teacher of your child's school has given permission for the school to take part, and your child's class will be involved in the study. Children in the class completed a computerised assessment (InCAS) at the beginning of the school year, which looks at reading, maths and attitudes to school. They will repeat this again at the end of the year. These assessments are used by hundreds of schools across the country and provide schools with information to help inform teaching. The data from this assessment will be passed on to Durham University along with some background information about children (i.e. date of birth, gender and unique pupil reference number) and used to look at how much children improve over the year.

All data collected will be treated confidentially and will only be used for research purposes. The data for your child's school will be analysed anonymously, together with data from other schools, and no individual pupils or schools will be named in any report. Pupil assessment data will be matched with the National Pupil Database for longer term follow-up and also shared with EEF data managers and the UK Data Archive for research purposes only.

The assessments are designed to be fun for the children to complete and we expect that they will enjoy doing them. I do hope that we have your support for this important study and that you are happy for your child's data to be used in this way. If you would prefer we did not use your child's data in this research study please contact us to let us know using the details below at any time during the project.

If you have any questions about the study, or what it involves for your child, or would like to opt out of the study, please contact Clare Collyer, Research Administrator, on 0191 334 4682, clare.collyer@cem.dur.ac.uk or Victoria Menzies, Project Researcher, on 0191 334 4177, victoria.menzies@cem.dur.ac.uk.

Yours faithfully,

A handwritten signature in black ink that reads 'Carole J. Torgerson'.

Carole Torgerson
Principal Investigator

Appendix D – Parent Consent (Pupil Focus Group)



School of Education
Durham University
Leazes Road
Durham
DH1 1TA

18th March 2015

Hallé SHINE on Manchester Evaluation by Durham University

Dear Parent/Carer

We are writing to the parents/carers of all pupils attending the Hallé SHINE on Manchester Saturday Schools. Durham University has been funded by the Education Endowment Foundation to evaluate the Hallé SHINE programme and to look at any improvements that could be made in the future. As part of our evaluation, researchers from the university will be visiting the project between now and July to observe a session and talk to the children attending and the staff running the programme.

We would like to hold a focus group with some of the participating children and would like your consent for your child to be included. The focus group will involve one or two researchers from the university meeting with a small group of pupils (4-5 children) and member of staff from the project. We will ask the children questions about their experience of attending SHINE. Participation in the focus group is voluntary and no individuals will be named in any report about the project. We expect the focus group to last for about 25 minutes and to take place during one of the Hallé SHINE sessions.

If you are happy for your child to be involved in a focus group please complete the tear off sheet below and return it to the Hallé SHINE on Manchester project next Saturday.

If you need further information about the evaluation or have any questions please contact the Evaluation Team through Clare Collyer, Project Administrator: clare.collyer@cem.dur.ac.uk; 0191 334 4682.

Yours faithfully

Professor Carole Torgerson
Principal Investigator

☐ I give consent for my child to be included in a focus group for the Hallé SHINE on Manchester evaluation to find out about the pupils' experiences.

Signed Date.....

Child's name

Appendix E: Analysis plan

1. Trial objectives

Primary objective

To compare performance in literacy (reading) between children who are allocated to take part in SHINE and those who are not allocated to take part in SHINE.

Secondary objective

To compare performance in literacy (spelling) and mathematics (general mathematics and mental arithmetic) between children who are allocated to take part in SHINE and those who are not allocated to take part in SHINE.

2. Sample size

The sample size calculation assumes there will be approximately 10 children per year group in each school taking part in the trial with an estimated intra-cluster correlation coefficient of 0.19 (which is based on a numeracy outcome) and the pre- post-test correlation of 0.74. At 5% significance level, 96 children are required to detect a difference of 0.58 with 80% power for the pilot trial. However, there are 170 children from 17 schools (3 hubs and 14 partner schools) recruited for the pilot trial.

3. Randomisation

The seventeen schools recruited for the pilot trial will be stratified by school type into three hubs schools, four hub 1 partner schools, seven hub 2 partner schools and three hub 3 partner schools. Since “school type” is the only important factor to be accounted for in the randomisation, the minimisation scheme reduces to stratified random allocation of schools in each stratum. To randomly allocate schools in each stratum to either year 5 or year 6 interventions, permuted block randomisation with a fixed allocation ratio will be used. The hub schools will be randomly allocated using a permuted block size of three with an allocation ratio of 2:1 in favour of the year 5 intervention group. The partner schools of hub 1 will be allocated to either year 5 or year 6 intervention groups using a permuted block size of four with an equal allocation ratio of 1:1. The partner schools of hub 2 will be randomly allocated to either year 5 or year 6 intervention groups using a combination of block sizes of 3 and 4 with allocation ratios of 2:1 and 1:1, respectively. The partner schools of hub 3 will be randomly allocated to either year 5 or year 6 intervention groups using a block size of 3 with an allocation ratio of 1:2 in favour of year 6 intervention group. Each block size will be randomly selected from its entire possible realisation based on the permutation of year 5 or year 6 intervention groups according to the specified allocation ratios. There are three possibilities using a block size of three with allocation ratios of 2:1 or 1:2 and six possibilities for a block size of 4 with an equal allocation ratio of 1:1. In summary, the randomisation will result in 9 schools randomised to year 5 intervention groups and 8 schools randomised to year 6 intervention groups. The year 5 intervention groups will be made up of two hub schools, two hubs 1 partner schools, four hub 2 partner schools and one hub 3 partner schools. The year 6 intervention groups will be made up of one hub school, two hub 1 partner schools,

three hub 2 partner schools and two hub 3 partner schools. The randomisation scheme will be implemented in R statistical software.

4. Outcomes

The standardised tests from the CEM InCAS system will be used to measure changes in attainment for all the children recruited for the pilot trial. Changes in attainment will be measured before and after the delivery of the intervention in all schools in all years for the pilot trial.

Primary outcome

The primary outcome measure for the pilot trial is literacy as measured by the children's reading score. The primary outcome will be assessed at baseline and post intervention.

Secondary outcomes

The secondary outcome measures are literacy as measured by the children's spelling score and mathematics (as measured by the children's general mathematics and mental arithmetic scores). General mathematics and mental arithmetic scores will be analysed separately. The secondary outcomes will be assessed at baseline and post intervention.

5. Analysis

All analyses will be conducted on an intention to treat basis, including all randomised children in the groups to which they were randomised. Analyses will be conducted in SAS version 9.3, using 2-sided significance tests at the 5% significance level.

Baseline Characteristics

Baseline characteristics (gender, age, free school meal status, baseline reading, baseline spelling, baseline general mathematics and baseline mental arithmetic scores) will be summarised by intervention group for each school and across all schools. Continuous variables (age, baseline literacy and baseline mathematics scores) will be summarised using descriptive statistics (n, mean, standard deviation, range and median). Categorical variables (gender and free school meal status) will be summarised using frequency counts and percentages by intervention group for each school and across the schools.

Trial completion

A CONSORT diagram will be used to present a summary of the flow of eligible children and their schools from recruitment through baseline assessment, randomisation, post intervention assessment and analysis. The number of children and schools included or excluded at each stage will be clearly stated and the reasons for exclusion will also be stated, where available.

Primary outcome

The primary outcome (reading score) will be analysed by class (year 5 and year 6) to test the null hypothesis that there is no difference in the average reading score between those randomised to SHINE and those not randomised to SHINE after accounting for baseline

reading score, gender, age, free school meal status, school type and potential clustering within schools.

A generalised linear model with Gaussian family distribution and identity link will be fitted to the data. The parameters of the model will be estimated by generalised estimating equation (GEE) with exchangeable working correlation for dependencies within schools in order to obtain robust standard errors for average differences between the intervention and control groups. This model will account for age, gender (1- female, 0-male), intervention groups (1-intervention, 0-control), school type and baseline reading scores. Model assumptions will be checked and if they are in doubt the data will be transformed prior to analysis or alternative parametric distribution that best reflects the data will be used. The average intervention effects, corresponding 95% confidence interval (CI) and associated p-value will be reported from the model. All model parameters will be reported.

Secondary outcomes

The three secondary outcomes: children's spelling, general mathematics and mental arithmetic scores will be analysed in the same manner as the primary outcome. Each model will account for age, gender (1- female, 0-male), intervention groups (1-intervention, 0-control), and school type as well as baseline scores of the corresponding secondary outcome. The average intervention effects, corresponding 95% confidence interval (CI) and associated p-value will be reported from the model. All model parameters will be reported.

6. Analytical Software

The data will be analysed with SAS^{®9} 9.3 Software. Specifically, the following SAS procedures will be used:

- *Proc data*: for data preparation
- *Proc means*: for producing descriptive statistics for continuous variables
- *Proc freq*: for producing descriptive statistics for gender
- *Proc Genmod*: for fitting GEE models for the primary and secondary outcomes.

Adetayo Kasim and Catherine Hewitt

290113